ABSTRACT

TZENG, SZ-TING. Human Centricity and Norm Awareness in Cognitive Systems. (Under the direction of Munindar P. Singh.)

Advancements in technology have seamlessly integrated Artificial Intelligence (AI) into our daily lives. Software engages in dynamic interactions with its environment, other software, and human beings, fostering a symbiotic relationship. The interconnectedness gives rise to a multiagent system (MAS) where humans and AI work together in synergy to attain shared objectives. Given the involvement of humans, AI systems must be able to reason over human behaviors, which are determined by a combination of internal attitudes and external factors. Incorporating human values and considering multiagent dynamics in decision-making would lead to a substantial improvement in the reliability and realism of AI systems. Besides aligning decisions with values, humans have a fundamental need to comprehend and place trust in the output of AI.

Another concern that arises from the growing size and dynamics of the MAS is adaptability. Social norms define acceptable group conduct and governing agent behaviors in MAS. Norms can arise through top-down imposition or bottom-up emergence. In both approaches, norms and the environment are subject to change over time. The capacity for adaptation in AI systems becomes crucial to minimize human intervention and effort in maintaining MAS.

In this dissertation, we aim to incorporate adaptability and explainability in AI systems by integrating normative MAS with human factors. Initially, we concentrate on elements that help to regulate human behaviors. Subsequently, we explore human factors associated with human needs. This research includes four components: emotions as sanctions, normative information from social signals, social value orientation, and value-aligned rationale. First, we introduce *Noe*, a framework that models the emotional responses of agents to the outcomes of interactions. Emotions, which are responses to internal or external events, can act as a positive or negative reinforcement mechanism for specific behaviors. Second, we introduce *Ness*, a framework that incorporates normative information from social signals to support norm emergence. In addition to sanctions, normative information from soft signals like hints and messages helps to regulate behaviors. Third, we present our *Fleur* framework, which incorporates the social value orientation concept. Social value orientation defines individuals' preferences over resource allocations between themselves and others. Aligning with social values enables AI to make ethical decisions and be responsible for human needs. Lastly, we describe *Exanna*, a framework that makes decisions and reveals information in rationales based on agents' values. Value-aligned

explanations ensure the AI system's decisions are consistent with human values and societal expectations.

Human Centricity and Norm Awareness in Cognitive Systems

by
Sz-Ting Tzeng

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Computer Science

Raleigh, North Carolina

2023

APPROVED BY:

Arnav Jhala	Min Chi
William Rand	Nirav Ajmeri

Munindar P. Singh
Chair of Advisory Committee

ACKNOWLEDGEMENTS

Thanks to family, advisor, committee, labmates, friends.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER

1

INTRODUCTION

Advancements in technology have seamlessly integrated Artificial Intelligence (AI) into our daily lives. For example, Netflix's recommendation system suggests videos based on users' preferences; virtual assistants on smart devices process and execute user requests in natural language. In contrast to the past, software is now unrestricted by confined and isolated environments. With cutting-edge technology, software interacts dynamically with its surroundings, other software, and human beings (Kafalı et al. 2016). This symbiotic interaction leads to the establishment of a multiagent system (MAS), where a synergistic relationship emerges between humans and AI. Given the involvement of humans, it becomes crucial to incorporate human factors while constructing contemporary AI systems. Specifically, AI systems must be able to reason over human behaviors, which are determined by both internal attitudes and external factors. AI would become more reliable and realistic by incorporating human values and considering the multiagent dynamics in decision-making.

In Attribution theory, internal attributions explain human behavior with a focus on the characteristics of a person (Gerace 2020). e.g., their personalities, abilities, and physical characteristics. On the contrary, external attributions stress environmental or situational factors. e.g., social influences and task difficulty. In the theory of basic human values, values characterize individuals and societies (Schwartz 2012). Values explain behaviors and attitudes on a motivational basis. Specifically, human values define an individual's intrinsic motivation and dominate how this individual thinks and evaluates everything. When a MAS becomes more interconnected, the complexity of interactions increases drastically, and it becomes hard to model all the possibilities. Basing on human values provides a solution to handle unexpected situations while aligning with human needs.

In addition to aligning decisions with values, humans have a fundamental need to comprehend and place trust in the output of AI. In other words, AI should be able to provide rationales for their decisions to be trusted by humans. However, explaining without considering individual differences may lead to information overload or privacy leaks among stakeholders.

Another concern that arises from the growing size and dynamics of the MAS is adaptability. Social norms play a significant role in MAS, defining acceptable group conduct and governing agent behavior (Savarimuthu and Cranefield 2011; Hollander and Wu 2011). A MAS incorporating norms that govern individual agents' behavior becomes a normative MAS. These norms elicit sanctions as responses to norm satisfaction or violation, e.g., penalties or rewards. Norms can arise through top-down imposition or bottom-up emergence (Morris-Martin et al. 2019). Top-down norms, e.g., laws and regulations, are costly and dictated by a central authority. Conversely, norms can also emerge from agent interactions in a bottom-up manner. In both approaches, norms and the environment are subject to change over time. The capacity for adaptation in AI systems becomes crucial to minimize human intervention and effort in maintaining MAS.

Sanctions, one form of social signals, coordinate and regulate agent behaviors in MAS. As humans evolved, social signals have emerged in the form of verbal messages or subtle hints, transmitting normative information.

While humans' decision-making includes internal and external attitudes, the other critical human factor in decision-making is emotions. Emotions, which are responses to internal or external events, can significantly impact decision-making and offer additional information in communication. Herbert Simon, a Nobel laureate, emphasized that general thinking and problem-solving must incorporate the influence of emotions (Simon 1967). Even more, emotions could be part of the norms themselves. Integrating both norms and emotions is essential for building explainable and trustworthy AI.

We aim to incorporate adaptability and explainability in AI systems by integrating normative multiagent systems (MAS) with human factors. Initially, we concentrate on elements that help to regulate human behaviors. Subsequently, we explore human factors associated with human needs. In this dissertation, we study methods to empower a normative MAS with the capability to adapt to dynamic environments and reason over human values.

1.1 Motivations

As AI systems increasingly involve humans in the decision-making process, there is a growing demand to consider human factors during their development. In this dissertation, we study four different human factors.

1.1.1 Emotion as sanctions

In multiagent systems, norms and sanctions are often used to regulate agent behaviors while maintaining their autonomy. However, sanctions in the real world are more subtle instead of harsh punishment. For instance, the sanctions could be trust update or emotional expression and might change one's behavior (Nardin et al. 2016; Bourgais et al. 2019). At the basic level of Emotions' Social Functions, emotions help individuals understand others' preferences, beliefs, and intentions and coordinate social interactions (Keltner and Haidt 1999).

Consider a pandemic scenario. During a pandemic, many stores limit the number of customers in stores at once to protect their customers. A side effect of this practice is the long queue outside the stores. While there is a social norm that people should line up to enter the stores, some can still jump the queue to get services in advance. Suppose those who violate the norms would feel guilty (self-directed emotion) and receive negative emotions from others (other-directed emotion). These felt emotions will enforce the norm in stores.

The above scenario demonstrates the necessity of incorporating emotions when studying norm enforcement.

1.1.2 Messages and hints as social signals

Social signals, as reactions to norm satisfaction or norm violation, provide natural drivers for norm emergence. When humans are evolved, social signals can be realized in three main ways: *sanction*, *tell*, and *hint*. Hints or emotions, as forms of non-verbal communication, are usually not considered in normative MAS. Normative information conveyed through a social signal also helps regulate MAS behaviors. In addition, hints may enable the inference of unobservable mental states (Wu et al. 2018; Wu and Schulz 2020).

With tell, agents communicate direct normative messages of approval or disapproval. An example of tell is verbal warning. An agent states clearly or indicates something may happen if someone does something. While messages provide clear normative information, hints also give subtle normative information for behavior. Upon receiving negative emotions after some actions, we can infer that our behaviors do not fit into others' expectations.

Consider the following example.

David notices Becka's suspicious symptoms and expresses some coldness near her. Upon perceiving David's negative attitude, Becka infers it is because she was sniffling near him in apparent violation of safety guidelines. Becka feels guilty about wandering out while being symptomatic. In addition, third parties who observe Becka's behavior and the actions of others may alter their behavior without directly having to be told.

The study of messages and hints as drivers of subtle social learning remains insufficient. Specifically, soft signals like hints have not been studied as drivers of norm emergence.

1.1.3 Social value orientation

While social norms regulate human behaviors, humans evaluate social norms based on human values and decide whether to comply or not. Social value orientation (SVO) is a psychological concept that describes individual differences in how people place value preferences along the dimensions of self and other. Consider there is a rare case scenario. During a pandemic, the authorities announce a mask-wearing regulation and claim that regulation would help avoid infecting others or being infected. Although Felix tests positive on the pandemic and prefers not to wear a mask, he also cares about others' health. If he stays in a room with another healthy person, Elliot, Felix will put the mask on.

While values may differ among individuals, a reliable AI system must take into account the values of its stakeholders to ensure making right decisions.

1.1.4 Decision and Rationale with values

Two key aspects are essential for AI systems to earn human trust and be interpretable. Firstly, the systems must align their decisions with human values at the micro level Liscio et al. (2023), focusing on the individual agent behaviors. While the macro level of values concerns the governance of MAS, the micro level ensures that decisions reflect individual human values. Secondly, agents should be capable of providing rationales for their decisions, enabling transparency and understanding behind their choices Winikoff et al. (2021); Ayci et al. (2023). Rationales serve as vital information for making decisions and can aid in resolving social conflicts. However, determining the appropriate extent of information an agent should provide raises crucial questions. One one side, overly detailed rationales might become convoluted and fail to persuade, resulting in information overload. One the other side, disclosing private information could leads to potential privacy breaches.

When humans are involved in the MAS, it becomes crucial for AI systems to make decisions that extend beyond mere physical gain and, instead, harmonize with human values. As values play a significant role in guiding motivations and steering decisions, rationales aligned with values best justify one's behaviors. In addition, values reflect different concerns in decision-making and conflict resolution among agents. The above issues demonstrate the necessity of building rationales based on agents' value.

1.2 Research Objective and Questions

Based on the aforementioned challenges, the research objective that we aim to achieve is to design a framework that incorporates human factors and operates in dynamic environments, ensuring the trustworthiness of AI systems.

In order to achieve our research objective, we seek to address the following questions.

RQ$_{\text{emotion}}$. How does modeling the emotional responses of agents to the outcomes of interactions affect the norm emergence?

RQ$_{\text{information}}$. How does considering soft signals such as hints and explicit normative messages in addition to sanctions influence norm emergence?

RQ$_{\text{SVO}}$. How do the preferences for others' rewards influence norm compliance?

RQ$_{\text{Rationale}}$. Do value-aligned rationales enrich the social experiences of agents?

1.3 Contributions

1.3.1 *Noe*: Enforcing social norm with emotions

To address RQ$_{\text{emotion}}$ that corresponded to Section 1.1.1, we refine the abstract normative emotional agent architecture (Argente et al. 2022) and investigate the interplay of norms and emotions. We propose a framework *Noe* based on BDI architecture (Bratman 1987), norm life-cycle (Savarimuthu and Cranefield 2011; Frantz and Pigozzi 2018; Argente et al. 2022), and emotion life-cycle (Alfonso Espinosa 2017; Marsella and Gratch 2009).

With *Noe*, agents can model emotional responses to the outcomes of norm satisfaction and violation.

1.3.2 *Ness*: Normative information from tell and hint

To address the problems in Section 1.1.2, we present an agent architecture integrating soft signals, such as messages and hints, to answer our first research question $\text{RQ}_{\text{information}}$. Our proposed framework *Ness* regards normative information from messages or hints as potential rewards which can shape behaviors. Including soft signals enables indirect social learning, which resembles human behaviors in the real world.

1.3.3 *Fleur*: Social value orientation for robust norm emergence

To tackle the challenges in Section 1.1.3, we develop *Fleur*, an agent framework that considers social value orientation, personal preferences, and social norms when making decisions. Specifically, SVO revises an individual's utility function by assigning different weights to itself and others. With *Fleur*, agents have the capability to make decisions aligning with their social value orientation.

1.3.4 *Exanna*: Decision and rationale with values

For the challenges in Section 1.1.4, we propose *Exanna*, a framework that incorporates values in decision-making, rationale generation, and reasoning over rationale. In addition, whereas other works focus on making agent decisions interpretable to humans, *Exanna* agents provide rationales to both agents and humans.

1.4 Organization

The dissertation is organized as follows.

Chapter 2 describes *Noe* and some related work. *Noe* shows how modeling the emotional responses of agents to the outcomes of interactions affect norm emergence and social welfare.

Chapter 3 introduces *Ness*, a framework that models normative information from social signals to support norm emergence. This chapter exhibits how agents with soft signals effectively avoid undesirable results and surpass baseline agents in delivering greater satisfaction despite requiring an equivalent amount of information.

Chapter 4 presents our *Fleur* framework to address RQ_{SVO}. *Fleur* incorporates the social value orientation, which provides agents with different preferences over resource allocations between themselves and others. This chapter shows how social value orientation enables better social experience and robust norm emergence.

Chapter 5 describes *Exanna*, which addresses $\mathrm{RQ}_{\mathrm{Rationale}}$ in simulated pandemic environments. This chapter demonstrates how value-aligned rationales enrich agents' social experiences.

Chapter 6 concludes this research and proposes possible future work.

CHAPTER

2

ENFORCING SOCIAL NORM WITH EMOTIONS

2.1 Introduction

Humans, in daily life, face many choices at many moments, and each selection brings positive and negative payoffs. In psychology, decision-making (Simon 1960) is a cognitive process that selects a belief or a series of actions based on values, preferences, and beliefs to achieve specific goals. Emotions, the responses to internal or external events or objects, can involve the decision-making process and provide extra information in communication (Keltner and Haidt 1999; Schwarz 2000). Social norms describe societal principles between agents in a multiagent system. While social norms regulate behaviors in society (Singh 2013; Savarimuthu and Cranefield 2011; Kafalı et al. 2020), humans and agents have the capacity to deviate from norms in certain contexts. For instance, people shake hands normally but deviate from this social norm during a pandemic. Chopra and Singh (Chopra and Singh 2016) describe how social protocols rely on a foundation of norms though they do not discuss how the appropriate norms emerge.

An agent that models the emotions of its users and other humans can potentially behave in a more realistic and trustworthy manner. The decision-making process for humans or agents involves evaluating possible consequences of available actions and choosing the action that maximizes the expected utility (Edwards 1954). Herbert Simon, one of the founders of AI, emphasized that general thinking and problem-solving must incorporate the influence of emotions (Simon 1967). Without considering emotions or other affective characteristics, such as personality or mood, some compliance seems irrational (Argente et al. 2022). Humans' compliance shows hints on rational planning over their objectives (Keltner and Haidt 1999). Including emotion or personality in normative reasoning makes these compliance behaviors explainable. Norms either are defined in a top-down manner or emerge in a bottom-up manner (Savarimuthu and Cranefield 2011; Morris-Martin et al. 2019). Works on norms include norm emergence based on the prior outcome of norms, automated run-time revision of sanctions (Dell'Anna et al. 2020), or considering various aspects during reasoning (Ajmeri et al. 2020, 2018). However, sanctions in the real world are often subtle instead of harsh punishments. For instance, sanctions could be trust updates or emotional expression and might change one's behavior (Nardin et al. 2016; Bourgais et al. 2019). Kalia et al. (Kalia et al. 2019) considered norm outcome with respect to emotions and trust and goals. Modeling and reasoning about emotions and other affective characteristics in an agent then become important in decision making and would help the agent enforce and internalize norms.

Accordingly, we propose *Noe*, an agent architecture that integrates decision-making with normative reasoning and emotions. We investigate the following research question.

RQ$_{\textbf{emotion}}$. How does modeling the emotional responses of agents to the outcomes of interactions affect norm emergence and social welfare in an agent society?

To address RQ$_{\text{emotion}}$, we refine the abstract normative emotional agent architecture (Argente et al. 2022) and investigate the interplay of norms and emotions. We propose a framework *Noe* based on BDI architecture (Rao and Georgeff 1991), norm life-cycle (Savarimuthu and Cranefield 2011; Frantz and Pigozzi 2018; Argente et al. 2022), and emotion life-cycle (Alfonso Espinosa 2017, pp. 62–64) (Marsella and Gratch 2009). To evaluate *Noe*, we design a simulation experiment with various agent societies. We investigate how norms emerge and how emotions in normative agents influence social welfare.

To make the problem tractable, we apply one social norm in our simulation and simplify the emotional expression to reduce the complexity. Specifically, our *Noe* agents process emotions by appraising norm outcomes. For the emotion model, we adopt the OCC model of emotions (Ortony et al. 1988) in which we consider both emotional valence and intensity and assume violation of norms yields negative emotions.

Organization. The rest of the paper is structured as follows. Section 2.2 discusses the relevant related works. Section 2.3 describes *Noe*, including the symbolic representation and the decision-making in *Noe*. Section 2.4 details the simulation experiments we conduct to evaluate *Noe* and describes the experimental results. Section 2.5 presents the conclusions and the future directions.

2.2 Related Works

Ortony et al. (1988) model emotions based on events, action, and objects. Marsella and Gratch (2009) proposed a computational model of emotion to model appraisal in perceptual, cognitive, and behavioral processes. Moerland et al. (2018) surveyed emotions in relation to reinforcement learning. Keltner and Haidt (1999) differentiate the functional approaches and research of emotions by four-level analysis: individual, dyadic, groups, and cultural. Briefly, emotions provide some information for agents or people to coordinate social interactions. We take inspiration from these works.

Savarimuthu and Cranefield (2011) proposed a life-cycles model for norms and discussed varied mechanisms of norm study. Broersen et al. (2001) introduced the so-called Beliefs-Obligations-Intentions-Desires (BOID) architecture on top of the Beliefs-Intentions-Desires (BDI) architecture (Rao and Georgeff 1991), which further include obligation and conflict resolution. de Lima et al. (2019) developed Gavel, an adaptive sanctioning enforcement framework, to choose appropriate sanctions based on different contexts. However, these works do not consider emotions in the decision-making process.

Argente et al. (2022) propose an abstract normative emotional agent architecture, which combines emotion model, normative model, and Belief-Desire-Intention (BDI) architecture. Argente et al. defined four types of relationships between emotions and norms: (1) emotion in the process of normative reasoning, (2) emotion generation with norm satisfaction or violation, (3) emotions as a way to enforce norms, (4) anticipation of emotions promotes internalization and compliance of social norms. Yet, Argente et al. do not validate the interplay between emotions and norms with their proposed architecture.

Bourgais et al. (2019) present an agent architecture that integrates cognition, emotions, emotion contagion, personality, norms, and social relations to simulate humans and ensure explainable behaviors. However, emotions are predefined and not generated via appraisal in this work.

von Scheve et al. (2006) consider emotion generation with norm satisfaction or violation. Specifically, an observer agent perceives the transgression of a norm of another, its strong negative emotions (e.g., contempt, disdain, detestation, or disgust) constitute negative sanctioning of the violator. The negative sanctioning then leads to negative emotions (e.g., shame, guilt, or embarrassment) in the violator. Besides, compliance with the social norms can stem from the fear of emotional-driven sanctions, which would lead to negative emotions in the violator. Such fear enforces social norms. Yet, emotions are not part of the decision-making process in this work.

2.3 *Noe*

We now describe the architecture, norm formal model, and decision-making.

2.3.1 Architecture

Noe integrates the BDI architecture (Rao and Georgeff 1991) with a normative model (Argente et al. 2022; Frantz and Pigozzi 2018; Savarimuthu and Cranefield 2011) and an emotional model (Alfonso Espinosa 2017; Marsella and Gratch 2009). A *Noe* agent assesses the environment, including other agents' expressed emotions, its cognitive mental states, and infer possible outcomes to make a decision. Figure 2.1 shows the three components of *Noe*.

The normative component of *Noe* includes the following processes:

- Identification: the agent recognize norms from its norm base based on its beliefs
- Instantiation: activate norms related to the agent
- Normative reasoning process: the reasoning process makes decisions based on the beliefs, current intention, self-directed emotions, other-directed emotions received from others, active norms, and how the norm satisfaction or violation influences the world and itself The *Noe* agents then update the intention based on the results of normative reasoning
- Norm fulfillment process: check if a norm has been fulfilled or violated based on the selected action. The compliance or violation of a norm will then trigger an elicit emotion event that will be appraised at the emotion component

The BDI component includes the following parts:

- Beliefs: form beliefs based on perceptions

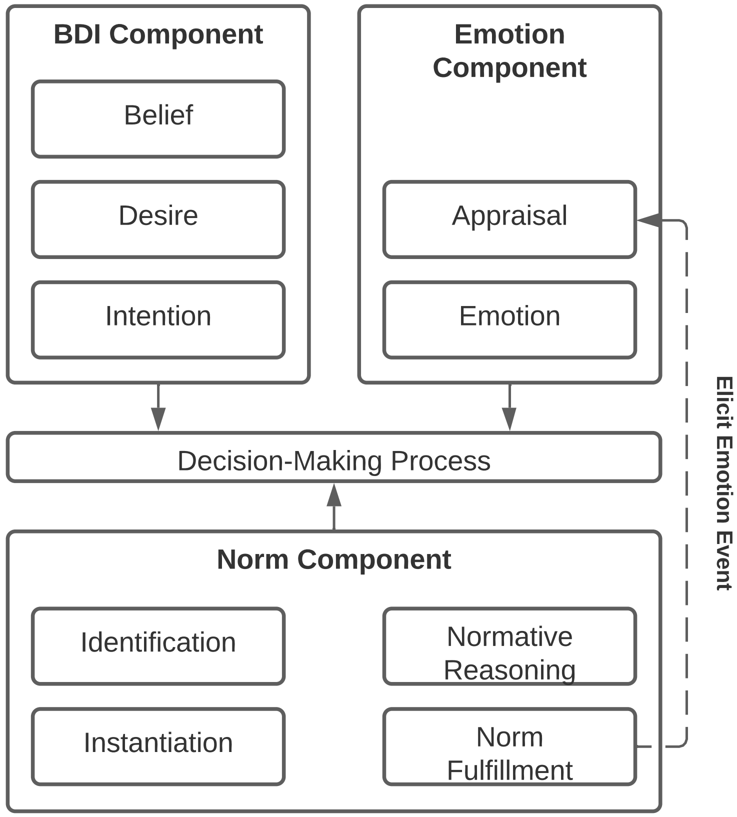


Figure 2.1: *Noe* architecture, representing and reasoning over beliefs, desires, intentions, emotions, and norms.

- Desires: generate desires based on the beliefs
- Intention: the highest priority of desires to achieve based on the beliefs
- Action: select action based on the current intention, emotions, possible outcomes, and the evaluation of violating or complying with norms, if any

The beliefs, desires, and intentions are mental states of *Noe* agents.

The emotional component includes the following processes:

- Appraisal: calculate the appraisal value based on the beliefs, desires, and norm satisfaction or norm violation. In this work, we consider only norm satisfaction or norm violation
- Emotion: generate emotion based on the appraisal values (Marsella and Gratch 2009)

Figure 2.2 illustrates the interactions between agents in our simulation scenario.

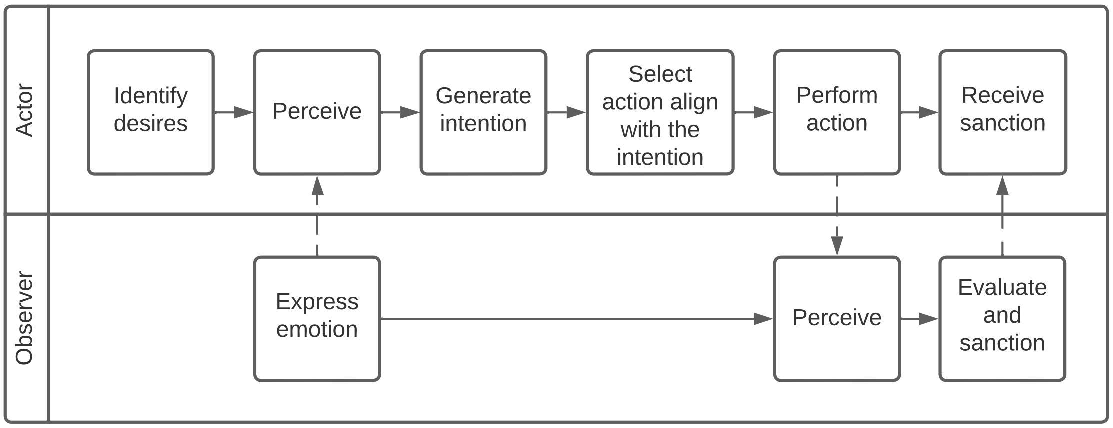


Figure 2.2: The interaction between *Noe* agents.

2.3.2 Norm formal model

Social norms describe the interactions between agents in a multiagent system. We adopt Singh's (Singh 2013) representation, where a social norm is formalized as $Norm(\text{subject}, \text{object}, \text{antecedent}, \text{consequent})$. In this representation, the subject and object represent agents, and the antecedent and consequent define conditions under which the norm is activated or satisfied, respectively. This representation describes a norm activated by the subject towards the object when the antecedent holds, and the consequent indicates if the norm was satisfied or violated.

Following Singh (Singh 2013), we consider three types of norms in *Noe*.

- Commitment (C): the subject commits to the object to bring out the consequence if the antecedent holds. Consider Alice and Bob are queuing up in a grocery store. Alice and Bob commit to keeping social distance during the pandemic, represented as $C(\text{Alice}, \text{Bob}, during = \text{pandemic}, \text{social distance})$.

- Prohibition (P): the object prohibits the subject from the consequence if the antecedent holds. Caleb, the grocery store manager, prohibits Bob from jumping the queue while lining up in that store, represented as $P(\text{Bob}, \text{Caleb}, when = \text{line up}; at = \text{grocery store}, \text{jump})$.

- Sanction (S): same as commitment or prohibition, yet the consequence would be the sanctions. Sanctions could be positive, negative, or neutral reactions to any norm satisfaction or violation (Nardin et al. 2016). If Bob breaks the queue, he receives negative sanctions from Alice, represented as $S(\text{Bob}, \text{Alice}, \text{jump}, \text{negative sanctions})$. Negative sanctions could be physical actions, e.g., scolding someone, or emotional expression, e.g., expressions of disdain, annoyance, or disgust.

To simulate the norm emergence and enforcement in human society, we include emotions into the decision-making process since, by nature, humans do not always act rationally in terms of utility theory. Here we formalize emotions with $E_i(target, intensity, decay)$ indicating agent a_i has emotion e toward the target with intensity and decay value. An example of the prohibition case would be, Bob would not jump the queue if Alice is angry, represented as $P(\text{Bob}, \text{Alice}, \text{Bob} \succ \text{Alice} \wedge E_{\text{Alice}} = \text{angry}, \text{jump})$.

We model the emotional response of agents with triggered emotions from norm satisfaction, or violation (Argente et al. 2022). Here we represent the elicited emotions with $Elem_{name}(A_{expect}, A_{real}, Em_1, Em_2) | Em_1, Em_2 \in E; A_{expect}, A_{real} \in A$ where A is a set of actions. E is a set of emotions, and Em_1 and Em_2 are the emotions triggered by norm satisfaction and violation accordingly. If the A_{expect} is equal to the A_{real}, a norm has been fulfilled, and Em_1 was elicited. $Ap(beliefs, desires, Elem)$ represents the appraisal function.

2.3.3 Decision-Making

Schwarz (Schwarz 2000) addresses the influence of moods and emotions at decision making and discusses the interplay of emotion, cognition, and decision making. Specifically, the aspects include pre-decision affect, post-decision affect, anticipated affect, and memories of past affect. In our model, we include the pre-decision affect into the decision-making process. With pre-decision affect, people recall information from memories that match their current affect (Schwarz 2000). For instance, people in a sad emotion or interacting with hostile people tend to overestimate adverse outcomes and events.

In our model, emotions serve as mental objects and an approach to sanctioning. We consider emotions as intrinsic rewards from agents' internal state in contrast to physical rewards from the environment. We adopt the OCC model of emotions (Ortony et al. 1988), in which we consider emotional valence and intensity. We formulate emotions with simple values where positive values indicate positive emotions and larger values indicate higher intensity. A mood is a general feeling and not a response to a specific event or stimulus compared to emotions. Therefore, we consider emotions but not mood. *Noe* agents' appraisal function considers norm satisfaction and violation only. The agents are aware of other agents' expressed emotions in the same place. In this work, we assume that agents express true and honest emotions and can correctly perceive the expressed emotions. In other words, felt emotions are equal to expressed emotions. Another assumption is that emotions are consistent with the notions of rational behavior.

Algorithm 1 displays the decision loop of our model. At the beginning of the simulation, all agents are initialized with certain desires, and during the run, an intention would be generated by prioritizing desires with the agent's beliefs. When choosing the next move with line 5 in Algorithm 1, the agent chooses the one with maximum utility from all available actions. Algorithm 2 details the action selection. The decision takes the agent's beliefs, current intention, and possible consequences into accounts. While norms are activated with the beliefs, the agent would further consider emotions and cost and possible consequences with norms at line 9 in Algorithm 2. For instance, if people violate some social norms, they may be isolated from society. Regarding the influence of emotions, people may overestimate the negative outcomes when they are in the negative emotion and tend to comply with the norms.

Algorithm 1: Decision loop of a *Noe* agent

---

1 Initialize one agent with its desires D;
2 **for** t=$1,T$ **do**
3   Observe the environment (including the expressed emotions from others E_{around}) and form beliefs b_t;
4   Generate intention I based on b_t and D;
5   a_t = ActionSelection(b_t, I, D);
6   Execute action a_t;
7   Elicit self-directed emotions E_{self} from agent itself based on if action a_t fulfills a norm;
8   Self-sanction with E_{self};
9   Observe the environment (including the performed actions a_{t_other} of other agents) and form beliefs b_{t+1};
10   Elicit other-directed emotions E_{other} for observer agents based on if action a_{t_other} fulfills a norm;
11   Sanction others with E_{other};
12 **end**

2.4 Evaluation

We evaluate *Noe* via a line-up environment where agents form queues to receive service. We detail the environment in Section 2.4.1.

Algorithm 2: Action selection

```
Input: beliefs b_t, intention I, desires D
Output: Action a_t
Function Action Selection:
    E_around ⊂ b_t;
    for each a in ACTIONS(b_t) do
        Activate norms N with beliefs b_t and a;
        if N = ∅ then
            a_t = MAX_a(RESULT(b_t, intention, a))
        else
            a_t = MAX_a(RESULT(b_t, intention, a, N) × amplifier(E_around) )
        end
    end
    return a_t
return
```

2.4.1 Line-up Environment

Figure 2.3 shows the line-up environment. We build this line-up environment using Mesa (Masad and Kazil 2015), a Python-based framework for building, analyzing, and visualizing agent-based models.

The line-up environment includes two shared locations—home and grocery stores. The agents move between home and grocery stores to get food. We consider one social norm in the line-up environment: agents are expected to line up to enter the grocery store. To simulate real human reactions to norm violations, we refer to a social psychology experiment (Milgram et al. 1986). In the line-up environment, we model defensive reactions of people in the queue as negative emotions toward those who jump the queue by barging in ahead of someone already in the queue. Conversely, people show positive emotions toward those who stay in the queue.

We initialize the agents with the following parameter values:

- Health (Integer value from 0–100): When the health value reaches zero, the agent is marked as *deceased* and unable to act. The health value decreases by 1 unit at each step.
- Deceased (Boolean: True or False): set as True when an agent runs out of health.
- Emotion (Integer value): simplified with numerical values where positive values indicate positive emotions and negative indicate negative emotions. The emotions come along with a duration. Default at 0.

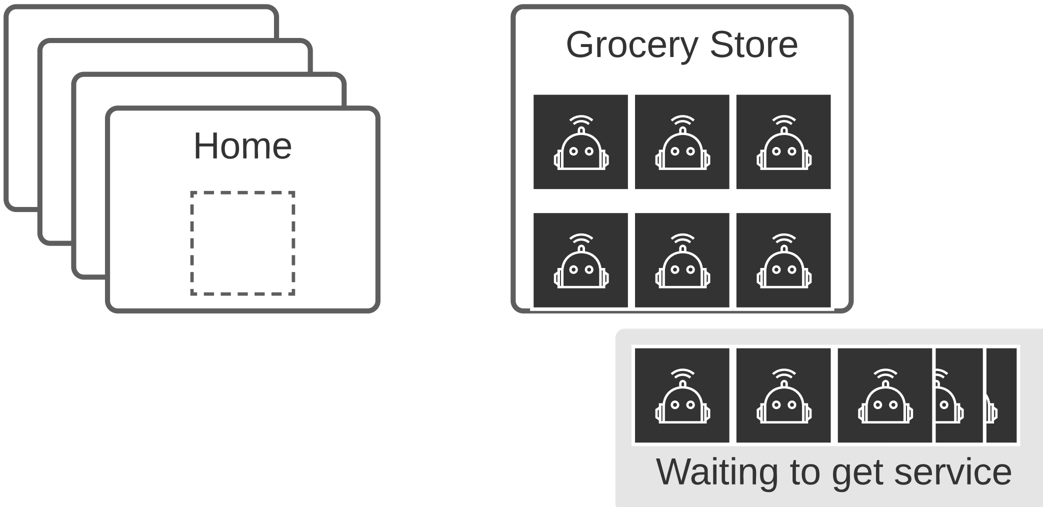


Figure 2.3: Simulation details. Agents move between their homes and the grocery store. The store has a capacity limit of eight customers at one time. As a result, other agents must line up outside the store to get service.

- Number of food packets owned (Integer value from 0–15): once obtained food from the stores, agents would be able to restore its health value via consuming food anywhere.
- Food expiration day (Integer value from 0–15): once the agent gets food packets, we update the expiration day with 15. The expiration day decreases by 1 unit at each step. Food expires once the expiration day reaches 0. Default at 0.
- Beliefs: the perceived and processed information from the world, including other agents' expressed emotions.
- Desires: desired states, including *have food* and *wandering*.
- Intention: the highest priority of desires to achieve at a specific time. When the agent's health is lower than the threshold, 80% of the health, the agent sets its intention as *get food*; otherwise, the agent sets its intention as *wandering*.

When an agent runs low on stock, it has a higher probability of moving to a grocery store. The grocery store can provide food packets to eight agents in one time step. While waiting in line to get food, the agent could either stay in the line or jump ahead in the line to get food in less time. Jumping the line may increase other agents' delay in getting food packets. Those who witness the violation would then cast negative emotions, further interpreted as anger or disdain, triggered by that behavior. To simplify the simulation, we presume the anticipated affects (Schwarz 2000) with: (1) receiving negative emotions triggers negative self-directed emotions such as shame and guilt; (2) complying with norms leads to positive or neutral emotions; (3) violating norms leads to negative or neutral emotions. The intensity of emotions triggered each time is fixed, but the values of emotions can add up. Each triggered emotion lasts 2 steps. At each step, the duration and intensity of emotion decrease by 1 as decay. A simple assumption here is that people in a bad mood would trigger stronger emotions in response to a non-ideal state. Note that at the beginning of the simulation, we initialize the agent society with health in normal distribution to avoid all agents having the same intention at the same time.

2.4.2 Agent Types

To answer our research question and evaluate *Noe*, we define three agent societies as baselines. We describe the agents societies below:

Obedient society. Agents in an obedient society always follow norms.

Anarchy society. Agents in an anarchy society jump lines when they cannot get food.

Sanctioning society. Agents in the sanctioning society jump lines considering the previous experience of satisfying or violating a norm. Agents sanction positively or negatively based on norm satisfaction or violations directly and comply with enforced norms.

***Noe* society.** Agents in the *Noe* society jump lines considering the previous experiences of satisfying or violating a norm, current emotional state of the other agents, current self emotional state, and estimated outcome of satisfying or violating a norm. *Noe* agents who observe norm satisfaction or violations would appraise the norm outcomes and trigger emotions to sanction the actor agent.

Table 2.1 summarizes the characteristics of the agents in the four societies.

2.4.3 Hypotheses and Metrics

To address our research question $\mathrm{RQ}_{\mathrm{emotion}}$ on emotions and norm emergence, we propose three hypotheses:

Table 2.1: Characteristics of the various agent societies.

Agent Type	Violation allowed	Sanctioning	Emotions involved
Obedient society	✘	✘	✘
Anarchy society	✔	✘	✘
Sanctioning society	✔	✔	✘
Noe society	✔	✔	✔

H_1 (Norm satisfaction): Norm satisfaction in *Noe* agent society is higher compared to the baseline agent societies.

H_2 (Social welfare): *Noe* agent society yields better social welfare compared to the baseline agent societies.

H_3 (Social experience): *Noe* agent society yields a better social experience compared to the baseline agent societies.

To evaluate H_1 on norm satisfaction, we compute one metric, M_1 (Cohesion): Percentage of norm satisfaction.

To evaluate H_2 on social welfare, we compute two metrics: (1) M_2 (Deceased): Cumulative number of agents deceased; (2) M_3 (Health): Average health of the agents. To evaluate H_3 on social experience, we compute one metric, M_4 (Waiting time): Average waiting time of agents in the queues.

To test the statistical significance of H_1, H_2, and H_3, we conduct the independent t-test and measure effect size with Glass's Δ for unrelated societies (Grissom and Kim 2012; Glass 1976). We adopt Cohen's (Cohen 1988, pp. 24–27) descriptors to interpret effect size where above 0.2, 0.5, 0.8 indicate small, medium, and large.

2.4.4 Experimental Setup

We run each simulation with 400 agents and queue size 80 for 3,000 steps. We choose a relatively small number of agents to reduce the simulation time while our results are stable for a more significant number of agents. The simulation stabilizes at about 1,500 steps, but we keep extended simulation steps to have more promising results. Table 2.2 lists the payoffs applied in our simulation.

We present the results with a moving average of 100 steps. We choose this size of running window to show the temporal behavior change in a small sequence of time. With a larger size, the running window may alleviate the behavior change. To minimize deviation from coincidence, we run each simulation with 10 iterations and compute the mean values.

Table 2.2: Payoff table.

Component	Type	Reward
Deceased	Extrinsic	–500
Norm compliance & positive emotion	Intrinsic	1
Norm violation & negative emotion	Intrinsic	–1

2.4.5 Experimental Results.

In this section, we describe the simulation results comparing the three baselines and *Noe* agents. Table 2.3 summarizes these results. Table 2.4 lists the value of Glass's Δ and *p*-values from the independent t-test.

According to Table 2.3 and Table 2.4, we see that *Noe* generate better cohesion and fewer deceased agents than baselines ($p < 0.01$; Glass's $\Delta > 0.8$). The null hypothesis corresponding to H_1 is rejected. Note that we do not consider the cohesion metric for the obedient agent society here since agents in the obedient society are always compliant. However, *Noe* also yields the worst social experience where the low waiting time is a desirable state ($p < 0.01$; Glass's $\Delta > 0.8$).

Table 2.3: Comparing *Noe* agent society with baseline agent societies on various metrics.

Agent Society	Cohesion	Deceased	Health	Waiting Time
Obedient	–	55.30	79.27	8.95
Anarchy	0.22	81.60	79.50	5.45
Sanctioning	0.88	169.30	86.26	2.55
Noe	**0.99**	**54.00**	79.00	8.95

Table 2.4: Statistical analysis.

Agent Society	Glass's Δ				*p*-value			
	Cohesion	Deceased	Waiting time	Health	Cohesion	Deceased	Waiting time	He
Obedient	0.19	0.65	0.01	0.18	0.32	< 0.01	0.98	0
Anarchy	102.43	3.10	40.82	0.21	< 0.01	< 0.01	< 0.01	0
Sanctioning	13.67	15.53	76.68	3.34	< 0.01	< 0.01	< 0.01	8
Noe	–	–	–	–	–	–	–	

H_1 Norm Satisfaction

Figure 2.4 displays the cohesion, the percentage of norm satisfaction, in the baseline agent societies and the *Noe* agent society. We find that the percentage of norm satisfaction in the *Noe* agent society, average at 99% and p-value < 0.01, is constantly higher than the sanctioning agent society, average at 88% and p-value < 0.01 and Glass's $\Delta > 0.8$. The sanctioning agent society learns to comply with the norm as time goes by. The *Noe* agent society does sanction as well. Yet, considering emotions and the possible outcome makes *Noe* agent society enforce the norm faster than the sanctioning agent society. Specifically, *Noe* agent society enforces the norm at about 100 steps while sanctioning agent society at 1,500 steps.

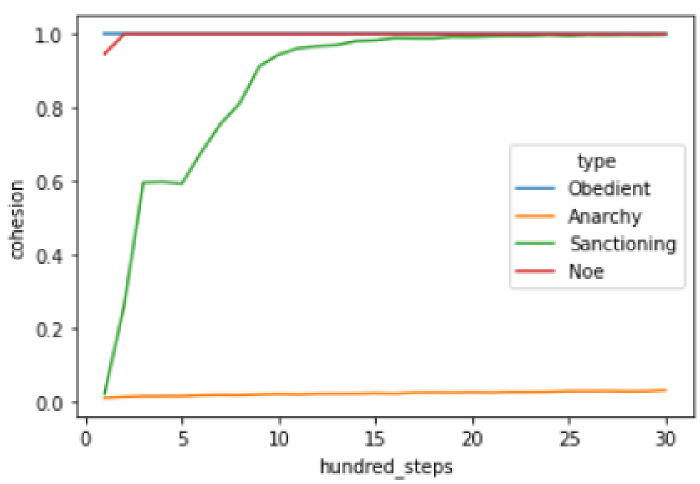


Figure 2.4: Simulation result: average cohesion. Comparing average cohesion (M_1) yielded by *Noe* and baseline agent societies.

H_2 Social Welfare

Figure 2.5 compares the average number of deceased in the obedient, anarchy, sanctioning, and *Noe* agent societies. Refer to Figure 2.4, sanctioning agent society learns the norm via positive and negative sanctioning from norm satisfaction and violation. However, the agents in that society do not consider the possible severe consequences and cause compliant agents to die in the queue. When the number of deceased reaches the threshold, the simulation stabilizes. Therefore, no more agent from the sanctioning agent society dies after the threshold. On the contrary, *Noe* agent society sanctions and considers possible outcomes of norm satisfaction and violation, therefore learning the norm and avoiding unacceptable consequences.

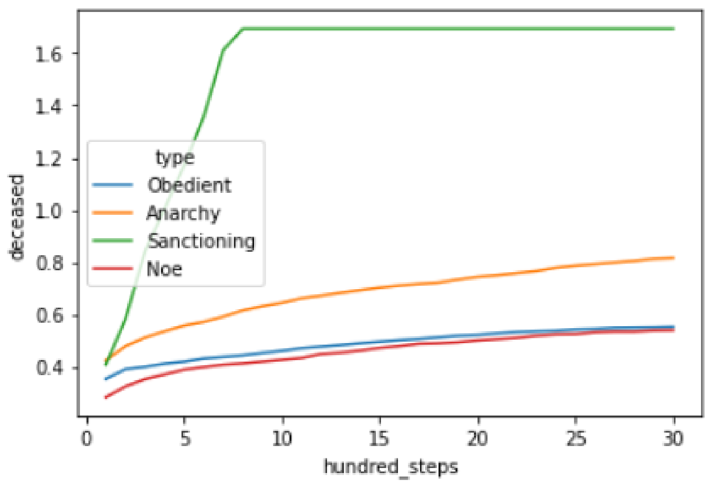


Figure 2.5: Simulation result: average number of deceased. Comparing average number of deceased (M_2) in *Noe* and baseline agent societies.

Figure 2.6 compares the average health of the agents in the obedient, anarchy, sanctioning, and *Noe* agent societies. The sanctioning agent society yields higher health State, with a mean at 86.26, but at the expense of more deaths. The rest of the agents then be able to remain in high health.

H_3 Social Experience

Figure 2.7 compares the average waiting time the agents spend in a queue at the grocery store in the obedient, anarchy, sanctioning, and *Noe* agent societies. The *Noe* agent society learns the norm fast and remains the same waiting time in the queue. However, some agents in the

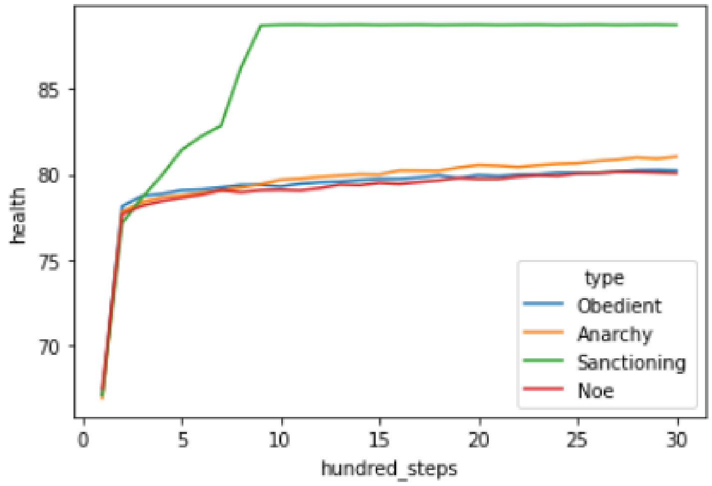


Figure 2.6: Simulation result: average health value. Comparing average health value (M_3) in *Noe* and baseline agent societies.

sanctioning agent society take advantage of those who learn norms faster than themselves. Therefore, many agents die during the learning process, and the simulation stabilizes. In Figure 2.7, the obedient agent society shares the same trend with *Noe* agent society since emotions enforce the line-up norm.

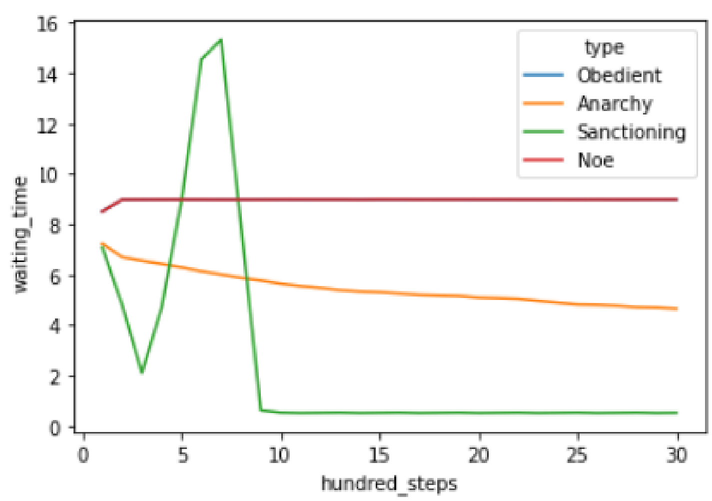


Figure 2.7: Simulation result: average waiting time of agents in queues. Comparing average waiting time (M_4) in *Noe* and baseline agent societies.

Combining the results for H_1 and H_2 and H_3, we note that while sanctioning enforces norms, a combination of sanctioning and emotions enforce norms better. Specifically, having emotions as amplifiers of outcomes yield higher norm satisfaction compared to our baselines. The results also indicate that, first, sanctioning agents that consider only norm violation or norm satisfaction may bring out worse social welfare compared to *Noe* that considers both norms and their consequences. Second, although *Noe* agents remain relatively high waiting time in the queues, the number of deceased is lower than the baselines. Note that the sudden drop of deceased number or increase of health value for sanctioning agents resulted from the stabilization of that society. Third, *Noe* agents stay in positive emotions during the simulation while sanctioning agents start from negative emotions and eventually achieve the expected behaviors.

2.5 Discussion and Conclusion

We present an agent architecture inspired by the norm life-cycle (Argente et al. 2022), BDI architecture (Rao and Georgeff 1991), and emotion life-cycle (Alfonso Espinosa 2017; Marsella and Gratch 2009) to investigate how emotions influence norm emergence and social welfare. We evaluate the proposed architecture via simulation experiments in an environment where agents queue up to receive service. Our simulations consider two characteristics of an agent society: sanctioning and emotions that participate in action selection and arise from evaluating selected action. The experiments show that incorporating emotions enables agents to cooperate better than those who do not.

In our agent architecture, we make an assumption that agents can recognize others' emotions. However, we acknowledge that emotion recognition is a challenging task (Barrett et al. 2019). Whereas recent works in AI have focused on emotion recognition through facial expressions and emotion recognition using wearables, it is worth noting that there is no agreement in modeling emotions in the psychology community (Barrett et al. 2019; Marín-Morales et al. 2018; Marsella et al. 2010).

Murukannaiah et al. (Murukannaiah et al. 2020) address many shortcomings of current approaches for AI ethics, including taking the value preferences of an agent's stakeholder and other agents' users, learning value preferences by observing the responses of other agents' users, and value-based negotiation. Incorporating these aspects in *Noe* is an interesting future direction.

As a future extension of current work, we plan to differentiate emotions in *Noe* instead of modeling emotions with emotion valences to provide more information for value preferences.

We also consider including a mix of personalities in future research to have different appraisal results. In this work, *Noe* agents are assumed to express true and honest emotions. However, emotions can also serve as a tool to influence, persuade, or deceive others in an adversarial context. It would be crucial to identify and model these contradictions while humans are in the loop.

CHAPTER

3

NORMATIVE INFORMATION FROM TELL AND HINT

3.1 Introduction

Social norms characterize collective and acceptable group conduct and regulate agent behavior. Norms may be imposed top-down (as legal norms are) or emerge bottom-up (when agents learn acceptable behaviors from each other) (Savarimuthu and Cranefield 2011). Our interest is in the latter. A norm emerges when the majority of agents in a society choose the same action (Morris-Martin et al. 2019; Savarimuthu and Cranefield 2011), sometimes termed concordance (Legro 1997).

We posit that the emergence of norms is driven by three kinds of social signaling between agents: (1) *Sanctions* or punishments or rewards (Nardin et al. 2016), (2) *Tell* or direct normative messages or explicit communications of approval or disapproval (Andrighetto et al. 2013), and (3) *Hint* or implicit signals conveying a positive or a negative attitude toward an observation.

Example 1 ***Sanction.*** *Becka is symptomatic with COVID-19. Alice meets Becka in a grocery store and notices Becka's symptoms. Alice omits to invite Becka from a party at Alice's home* *that Becka knows of.*

Example 2 ***Tell.*** *Charlie notices Becka's suspicious symptoms and begins to worry about his safety. He tells Becka that wandering around in public while symptomatic is not right.*

Example 3 ***Hint.*** *David notices Becka's suspicious symptoms and expresses some coldness near her. Upon perceiving David's negative attitude, Becka infers it is because she was sniffling near him in apparent violation of safety guidelines. Becka feels guilty about wandering out while being symptomatic.*

In each case, via a social signal, Becka learns that her behavior was inappropriate. Further, third parties who observe Becka's behavior and the actions of others may alter their behavior without directly having to be told.

Messages and hints drive subtle forms of social learning, as in human societies, but have not been adequately studied. Soft signals such as a hint, in particular, have not been studied as drivers of norm emergence. We investigate the following research question.

$\text{RQ}_{\text{information}}$. How does considering soft signals such as hints and explicit normative messages in addition to sanctions influence norm emergence?

To address $\text{RQ}_{\text{information}}$, we define two expressions of normative information: explicit normative message (Andrighetto et al. 2013) and implicit hint as information.

The contributions of this work are as follows. We propose *Ness* (for *Norm Emergence through Social Signals*), a framework that accommodates norms imposed top-down and enables norm emergence. *Ness* includes normative information from three types of social signals; that information facilitates social learning.

During Covid-19, abundant research is investigating the effects of interventions against the spread of the virus. However, little research considers policy violations, which are the essential drivers of a pandemic. Modeling social signals with a framework enables a more realistic simulation of individuals' decisions, e.g., obedience or noncompliance to interventions against the spread of the pandemic.

We evaluate *Ness* via simulation experiments with a pandemic scenario. We consider agent societies with a combination of three signal characteristics: sanction, tell or direct messaging, and hint. Our results demonstrate that normative information from hint and direct messaging enables faster norm emergence, avoids undesirable consequences, and yields higher satisfaction.

The rest of this paper is structured as follows. Section 3.2 introduces key concepts of *Ness* and describes how agents' decision-making work in *Ness*. Section 3.3 details the pandemic simulation environment we create to evaluate *Ness*. Section 3.4 presents results from our simulation experiments. Section 3.5 concludes other relevant research. Section 3.6 concludes with a summary of our findings, limitations, threats to validity, and future directions.

3.2 *Ness*

A *Ness* agent selects actions considering their goals, environmental norms, and social signals (Argente et al. 2022; Frantz and Pigozzi 2018; Savarimuthu and Cranefield 2011; Singh 1994; Marsella and Gratch 2009).

A *Ness* agent learns from observations. Following Example 2, on receiving Charlie's message, Becka learns that she may be reported to local authorities if she does not self-quarantine. In Example 3, Becka may have misread David's coldness as directed at her.

With recognized norms, a *Ness* agent activates those norms related to itself based on its knowledge. The normative reasoning process enables *Ness* agents to reason over the possible outcome of norm compliance or violation. After executing a chosen action, the agent checks if a norm has been fulfilled or violated. The compliance and violation of norms then trigger social signals.

3.2.1 Key Concepts

We now introduce the key concepts in *Ness*.

Goal is a condition that an agent wants to achieve. The outcome of a goal has a binary value, achieved or not, after performing a selected action.

Norm defines the relationship between an agent on whom the norm is focused and an agent with respect to whom the norm arises. An agent can invest effort on its norm or has its freedom curtailed by the object. A norm can either be satisfied or violated when the consequent holds or not, respectively.

Sanction refers to a positive, negative, or neutral reaction directed from one agent toward another. A sanction is typically in response to a norm satisfaction or norm violation.

Tell or normative information expression specifies the cause and the effect. For example, Charlie's specification in Example 2 includes whether a norm is satisfied or violated and why.

Hint describes a state that agent has and expresses toward another agent. We model hints as subtle soft signals triggered by norm satisfaction or violation.

Reward Shaping With reward shaping (Marom and Rosman 2018), agents are provided with "shaping" rewards (in addition to those from the environment) to move towards the goal or to encourage taking some actions in some set of states. Shaping rewards can be potential rewards from exterior knowledge or information, represented as potential functions, that guide an agent to avoid or prioritize specific behaviors. Here, we consider normative information from messages or hints as advice on potential soft sanctions with different levels of certainty. That is, signals—*tell* and *hint*—are inferred as positive or negative rewards to encourage or discourage taking specific actions.

3.2.2 Decision-Making

An agent's behaviors include acting to maximize its payoff, and giving social signals.

Action selection An agent selects an action that satisfies its goal and maximizes its actual and possible payoff. In the pandemic examples in Section 3.1, Becka decides whether to go to the grocery store depending on her goals and her understanding of norms.

Social signal expression An agent observes other agents' behaviors and expresses social signals — sanctions, direct messages, or hints. Specifically, an agent evaluates whether the observed behaviors conflicts with norms. In Example 1, Alice on noticing Becka's symptoms sanctions Becka based on her understanding of a healthcare guidance of staying at home when symptomatic. In Example 2, Charlie sends direct messages to correct Becka's behavior. In Example 3, David's coldness toward Becka shows his disapproval. On observing these hints, Becka and others nearby may relate David's coldness to Becka's behavior of deviating from a norm.

Reward Shaping A message or hint serves as a look-ahead advice on what will happen after a specific action. A shaping reward can be defined as $r' = r + F$ where r is the original reward function, and F is the shaping reward function. With messages or hints, F defines the difference of potential values.

$$F(s,a,s',a') = \gamma\Phi(s',a')\kappa - \Phi(s,a) \tag{3.1}$$

where Φ is a potential function that gives hints on states. κ defines the certainty of potential reward from the knowledge or information.

3.3 Simulation

We evaluate *Ness* via a simulated pandemic scenario where how agents behave influences the spread of a pandemic. We built our pandemic environment using Mesa (Masad and Kazil 2015), a Python-based simulation framework. Our focus is not to model the realism of pandemic spreading but to investigate how social signals influence norms. Agents in the simulation use reinforcement learning to learn the relationship between objectives and normative behaviors.

3.3.1 Pandemic Scenario

In the simulated environment, an agent moves between four places (home, park, grocery store, and hospital. Each agent's home is different). An agent has a goal to rest, hike, shop, be_-vaccinated, and selects actions from {stay_home, visit_park, visit_grocery store, visit_hospital}. Any two agents present at the same place may interact with equal probability at each step. Agents perceive each other's signals, and all expressed signals are genuine and honest.

We adapt the Susceptible-Exposed-Infected-Recovered-Vaccinated (SEIRV) (Yang and Wang 2020; Annas et al. 2020) disease model to simulate the pandemic. An agent becomes infected with probability p upon interacting with an infected agent. We base the probabilities on how COVID-19 evolves on a study in Italy (Poletti et al. 2020). Section 3.3.2 provides details about the disease model.

At each step, an agent observes its environment and moves accordingly. After each agent moves, the agents may signal (sanction, tell, or hint) each other based on their and others' behaviors. An agent who witnesses another agent being signalled can learn from it. The elements based on which an agent learns where to move include death, goal satisfaction, sanctions, messages, hints, and norm satisfaction or violation. These are modelled as elements for reward (Section 3.3.7).

3.3.2 Disease Model

After the breakout of COVID-19, vaccines provide substantial protection against severe illness and hospitalization. However, vaccines may not always provide complete protection. To maintain the dynamism, we include the effectiveness of vaccines in the Susceptible-Exposed-Infected-Recovered-Vaccinated (SEIRV) model (Yang and Wang 2020; Annas et al. 2020). Our adapted SEIRV model separates the deceased compartment from the recovered (R), so the state transitions are complete and intuitive. As shown in Figure 3.1, each agent starts with being susceptible (S) to the virus, can be exposed (E) to the virus when that agent has any contact with an infected agent. Susceptible and exposed and recovered are healthy agents. Being exposed to the virus, an agent can become Infectious (I). We further define Infectious as having three subclasses: asymptomatic, mildly symptomatic, and critical symptomatic. After some immune response, an agent becomes either recovered or deceased from the virus. Those who recover from the virus gain antibodies that protect them from the same virus. The other way to gain protection is vaccination. Once people obtain antibodies from vaccines or recovery, they have fewer opportunities to be infected.

We base the probabilities of how COVID-19 evolves on a study in Italy (Poletti et al. 2020). According to the quantifying research, among the close contacts, 51.5% were infected. However, the B.1.617.2 (Delta) variant is 50% more contagious than the original strain of SARS-CoV-2. Morbidity and mortality weekly reports (Thompson et al. 2021) show that the effectiveness of mRNA vaccine with full immunization was 90% against COVID-19 infections regardless of symptom status. With different vaccines, the effectiveness with full vaccination ranged from 67% to 93.7% among persons with the Delta variant (Lopez Bernal et al. 2021). We set the infection probability to 80% and the effectiveness of vaccination at 50% to represent a more infectious variant and speed up the simulation. Apart from vaccination, we set the probability of the symptoms to progress as Figures 3.1 and 3.2. The intuition is that each infected person provides an opportunity for the symptoms to progress to the next phase or recover. We assume infected agents that stay at home have a better recovery rate than those hanging outside based on common sense. With complete vaccination, the probability of the disease evolving is halved. Table 3.2 details the transition probability.

We have a greater tendency to sanction those who severely violate norms in the real world. Therefore, we set a 50% probability to sanction those who appear mild symptoms and an 80% possibility to sanction those who appear critically symptomatic but are not quarantining in our simulation. Since humans have partial observability of others in the real world, we include uncertainty of observing other health states (Table 3.1). Someone who is sniffling has some probability of being perceived as symptomatic.

Table 3.1: Uncertainty on observing others' state.

Actual \ Belief	Healthy	Symptomatic	Critical Illness
Healthy	0.8	0.1	0.1
Asymptomatic	0.5	0.5	0.0
Symptomatic	0.3	0.6	0.1
Critical Illness	0.1	0.3	0.6

Table 3.2: State transition of disease with probability. Note that the transition for healthy agents applies when coming in contact with those who are infected.

s_t \ s_{t+1}	Healthy	Asymptomatic	Mild	Critical	Deceased
Healthy	$1 - 0.8\alpha$	0.8α	0	0	0
Asymptomatic	0.2β	$1 - 0.36\alpha - 0.2\beta$	0.36α	0	0
Symptomatic	0.1β	0	$1 - 0.01\alpha - 0.1\beta$	0.01α	0
Critical	0.05β	0	0	$1 - 0.2\alpha - 0.05\beta$	0.2α

3.3.3 Social Norm

We initialize the environment with a social norm that healthy agents prohibit infected agents from staying in a public space. In other words, healthy agents may force infected agents to quarantine. We frame the norm as commitment as below.

{antecedent={observed health=[SYMPTOMATIC_INFECTED, CRITICAL]}, conse

When the antecedent and consequent both hold, the commitment is fulfilled, and a sanction is given to the agent. The sanction is a numerical reward to the agent who fulfills the commitment. When an agent receives normative information from others indicating or stating this commitment, it considers the sanction a potential reward Φ. According to the signal type, the agent constructs the shaping reward.

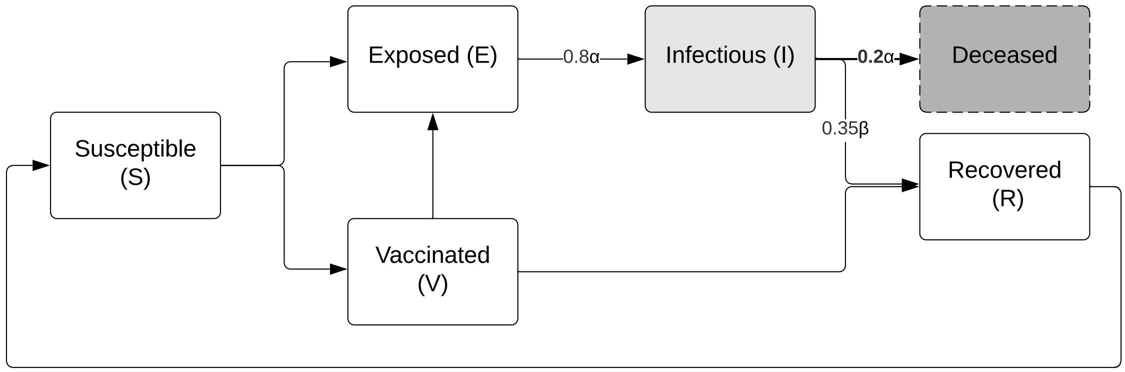


Figure 3.1: Disease model with state transition probabilities adapted from SEIRV disease model (Yang and Wang 2020; Annas et al. 2020). Agents may progress between compartments with the given probability. We set α to 0.5 for vaccinated agents and to 1.0 for unvaccinated agents; $\beta = 1.0$ for agents not staying home and $\beta = 2.0$ for agents staying home. The probability of remaining in the state is $1-$ the probability of evolving to the next state.

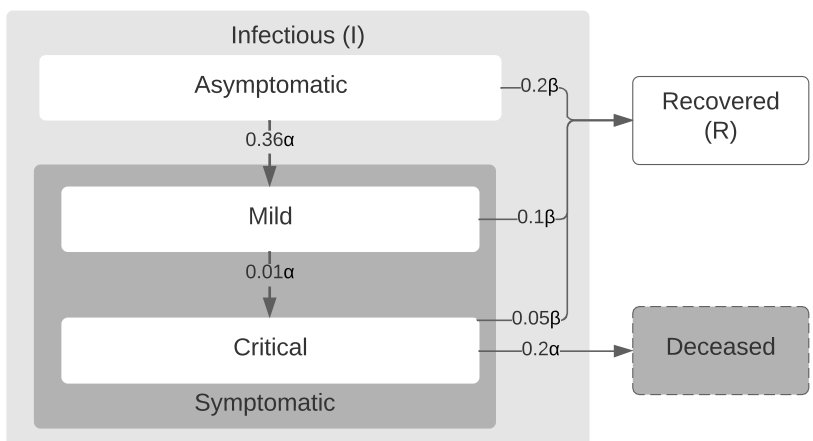


Figure 3.2: State transitions between infectious, recovered, and deceased. Infectious include three subclasses: asymptomatic, mildly symptomatic, and critically symptomatic. Numbers on the edges show the probabilities of transition. Hence, $\alpha = 0.5$ for vaccinated agents and $\alpha = 1.0$ for unvaccinated agents. And, $\beta = 1.0$ for agents not staying home and $\beta = 2.0$ for agents staying home. The probability of remaining in the state is $1-$ the probability of evolving to the next state.

3.3.4 Types of Agent Societies

We consider various agent societies with a combination of three social signaling characteristics: sanctioning, telling or direct messaging, and providing hint.

Baseline 1: PRIMITIVE society

Agents in a primitive society do not have shared desirable behaviors approved by most agents. In addition, agents apply no social signals. They act based on payoff from goal satisfaction and survival.

Baseline 2: SANCTION society

Agents obtain negative sanctions for violating a social norm, as in Example 1. In the simulation environment, infected agents who enter a public space may be forced to quarantine at home by healthy agents.

Baseline 3: HINT society

With some social norm in mind, agents who violate or satisfy the social norm receive a signal from others. These agents may also experience guilt or pleasure based on norm violation or satisfaction. This society is a variant of the *Ness* society but without the shaping rewards part. In the simulation environment, infected agents who wander in a public space may feel bad about it. They may also receive hints of norm violation from healthy agents and may be forced to quarantine at home by healthy agents.

Baseline 4: TELL society

Agents learn social norms and convey normative messages upon witnessing a norm violation. The normative message is adapted from (Andrighetto et al. 2013) and includes what sanctions an agent will receive if it violates a norm. An example of this society is Example 2. In the simulation environment, healthy agents interacting with infected agents in a public space convey the social norm of staying away from public spaces to the infected agents. Also, infected agents in a public space may be forced to quarantine by healthy agents.

***Ness*: Society combining social signals**

A society with our proposed implemented agents. *Ness* agents learn norms from social signals of sanction, tell and hint. In the simulation environment, infected agents who wander in public spaces receive hints from healthy agents via which they infer the normative information about staying away from healthy agents. This information from the hint signal provides shaping rewards to agents to learn norms. *Ness* agents may also experience pleasure or guilt based on norm violation or satisfaction. In addition, similar to SANCTION, HINT, and TELL societies, infected agents in a public space may be forced to quarantine at home by healthy agents.

3.3.5 Metrics

To compare the baseline societies with *Ness*, we compute these measures:
$\mathbf{M_{Healthy}}$. The percentage of healthy agents.
$\mathbf{M_{Infected}}$. The percentage of agents who are infected.
$\mathbf{M_{Deceased}}$. The percentage of deceased agents.
$\mathbf{M_{Infections}}$. The total number of infections in societies.
$\mathbf{M_{Vaccinated}}$. The percentage of vaccinated agents.
$\mathbf{M_{Isolation}}$. The percentage of self-isolation among infected.
$\mathbf{M_{Forced\ quarantine}}$. Number of agents who are forced to quarantine at home. This measure maps to the signal of sanction.
$\mathbf{M_{Goal}}$. The average goal satisfaction among agents.

We compute $\mathrm{M_{Healthy}}$, $\mathrm{M_{Infected}}$, $\mathrm{M_{Deceased}}$, $\mathrm{M_{Infections}}$, $\mathrm{M_{Vaccinated}}$, and $\mathrm{M_{Goal}}$ to identify the consequences of agents' behaviors or norm emergence if any. Moreover, these measures provide hints on why specific norms emerge. $\mathrm{M_{Isolation}}$ and $\mathrm{M_{Forced\ quarantine}}$ disclose agents' behaviors on the emerging norms.

A norm emerges when the proportion of agents following the same behavior exceeds a threshold. We consider 90% as the threshold (Delgado 2002).

3.3.6 Hypotheses

We evaluate three hypotheses respectively corresponding to specific metrics.
$\mathbf{H_{Disease\ control}}$. Agent societies considering hints have better control over disease spread than the societies that do not consider hints. We compare $\mathrm{M_{Healthy}}$, $\mathrm{M_{Infected}}$, $\mathrm{M_{Deceased}}$, $\mathrm{M_{Infections}}$, and $\mathrm{M_{Vaccination}}$.

$\mathbf{H_{Isolation}}$. The proportion of isolation is higher in agent societies considering hints than those that do not consider hints. We compare $M_{Isolation}$ and $M_{Forced\ quarantine}$ and $M_{Infected}$.

$\mathbf{H_{Goal}}$. Agents in *Ness* society have higher goal satisfaction than other baseline societies. We compare M_{Goal}.

To test the statistical significance of our hypotheses, we conduct the independent t-test and measure effect size with Glass' Δ since the standard deviations are different between societies (Glass 1976; Grissom and Kim 2012). We adopt Cohen's (Cohen 1988) descriptors to interpret effect size where 0.2 indicate small, 0.5 indicate medium, and 0.8 indicate large effect. An effect size less than 0.2 indicates that the difference is negligible.

3.3.7 Experimental Setup

Here we describe our simulation settings. Table 3.3 lists the elements of reward function, including extrinsic rewards from the environment and intrinsic rewards from agents' internal state.

We consider 100 agents (of whom 30 are infected initially) with the simulated world lasting for 2,000 steps, each interpreted as one day, which is adequate for our study since each society stabilizes within 1,500 steps. We train our agents for 100,000 steps, and average results over 20 runs.

For data efficiency, we apply policy parameter sharing (Gupta et al. 2017) based on the assumption of the bystander. Although policy sharing makes agents homogeneous, the agents make different observations and behave accordingly.

We consider normative information as shaping rewards that are part of intrinsic rewards. Specifically, we incorporate knowledge of being punished in the future from messages or hints into our simulation. Table 3.3 lists elements based on *Ness* agents appraises their states.

Information Balance

As messages and hints provide additional normative information to learn, we keep the expected payoff at the same level to balance the amount of information agents receive from combination of signals.

Intuitively, agents can learn more efficiently if more learning channels are provided. To achieve comparability, we balance the information an agent can access by adjusting the expected payoff. Table 3.4 lists the signal probability we apply for each agent society. The sanction defines the probability that an agent receives sanctions with a weight of 1.0 from its neighbors. Tell indicates the probability that an agent receives messages and considers them as shaping

Table 3.3: Reward function. Hard sanctioning means forced quarantine. Extrinsic rewards come from the environment, and intrinsic rewards come from agents' internal states. Note that the norm satisfaction or violation evaluates the action and the perceived health state of others instead of the actual health state.

Component	Type	Reward
Deceased	Extrinsic	–2
Sanctioning	Extrinsic	–1
Goal satisfaction	Intrinsic	+1
Goal violation	Intrinsic	–1
Norm satisfaction (self)	Intrinsic	+1
Norm violation (self)	Intrinsic	–1
Norm satisfaction (other)	Extrinsic	+1
Norm violation (other)	Extrinsic	–1

rewards with a weight of 0.5 from its neighbors. Hint specifies the probability that an agent receives implicit signals conveying a positive or a negative attitude toward its observation with a weight of 1.0 from its neighbors. Hint w/ shaping rewards defines the probability that an agent receives implicit signals of positive or negative attitudes toward its observation with a weight of 1.0 from its neighbors. Furthermore, the agent infers the signals as positive or negative rewards with a weight of 0.3 to encourage or discourage specific behaviors.

Table 3.4: Signal distribution. In each society, agents have some probability of sending the combination of the three social signals. To balance the amount of information, we manage and keep the expected payoff from these signals at the same level.

Societies: Signals	Sanction	Tell	Hint	Hint w/ shaping reward
PRIMITIVE	0%	0%	0%	0%
SANCTION	38%	0%	0%	0%
TELL	20%	36%	0%	0%
HINT	20%	0%	12%	0%
Ness	20%	0%	0%	10%

Reinforcement Learning Parameters for Social Signals

Ness and baseline agents employ Q-Learning (Watkins and Dayan 1992) to learn norms. Q-Learning is a model-free reinforcement learning algorithm that learns from trial and error with given rewards or penalties. Q-Learning algorithm computes the action-state value $Q(s,a)$ (Q value), which indicates the expected and cumulative rewards for each state and action. By approximating the value of an action for a given state, the Q-Learning algorithm finds the optimal policy. The Q function that computes Q values with the weighted average of the old value and the new information::

$$Q'(s_t,a_t) = Q(s_t,a_t) + \alpha * (r_t + \gamma \max_{a'} Q(s_{t+1},a) - Q(s_t,a_t)) \tag{3.2}$$

where $Q'(s_t,a_t)$ represents the updated cumulative value of $Q(s_t,a_t)$ after performing action a at time t. In this equation, α indicates the learning rate, and γ defines the reward discount rate.

Messages give precise causality between claimed behaviors and possible sanctions. On the contrary, hints provide subtle normative information for behaviors, which requires further inference. While hints and messages provide different levels of certainty of possible sanctions, we model normative information with approving and disapproving attitudes as shaping rewards. Specifically, we associate approving and disapproving attitudes with norm satisfaction and violation. The potential rewards are calculated from the signal based on the signal type. We set κ as 0.3 for hints and κ as 0.5 for messages where κ is the certainty of possible sanctions from normative information. Table 3.5 lists the parameters in our simulation.

Table 3.5: Hyperparameters.

Parameter	Value	Comment
Learning rate α	0.0010	
Discount factor γ	0.9	
Simulation step per action	1	
Infection %	0.3	The default fraction of infected agents in a society
Certainty of potential reward κ	0.3; 0.5	0.3 and 0.5 are certainty of possible sanctions from normative information through hints and messages, respectively.

3.4 Experimental Results

We now discuss the results for our research question $RQ_{\text{information}}$. Table 3.6 summarizes the simulation results and the corresponding statistical analysis for $RQ_{\text{information}}$. The metric row for each hypothesis in the table shows the numeric value of the metric after convergence of the simulation runs. All numeric values in Table 3.6 are rounded to 4 decimal places.

3.4.1 $H_{\text{Disease control}}$

To evaluate $H_{\text{Disease control}}$, we measure the proportion of healthy (M_{Healthy}), infectious (M_{Infected}), and deceased (M_{Deceased}) agents. In addition, we track the total number of infections ($M_{\text{Infections}}$) and the vaccination rate ($M_{\text{Vaccinated}}$) in each society. Infectious agents include those who are asymptomatic, mildly symptomatic, and critical symptomatic.

Figure 3.3 shows the comparison of M_{Healthy}, M_{Infected}, M_{Deceased}, and $M_{\text{Vaccinated}}$ in agent societies. Results in Figure 3.3 start from step one while the simulation starts with a 30% infection rate in a society. We observe that, first, the *Ness* society has lower infected agents (0.1567) than PRIMITIVE (12.6231), SANCTION (2.6343), HINT (0.4109) , and TELL (4.2049) societies. The differences in the results are statistically significant with a large effect ($p < 0.001$; Glass' $\Delta > 0.8$) for PRIMITIVE and HINT societies and with a small effect ($p < 0.001$; Glass' $\Delta > 0.2$) for SANCTION and TELL societies.

Second, the *Ness* society has the most healthy agents (98.7499) than PRIMITIVE (46.3426), SANCTION (77.601675), HINT (96.6218), and TELL (65.082) societies. These differences in the results are statistically significant with a large effect ($p < 0.001$; Glass' $\Delta > 0.8$).

Third, the *Ness* society has lower deceased agents (1.0934) than PRIMITIVE (41.0342), SANCTION (19.7640), HINT (2.9674), and TELL (30.7130) societies. The differences in the results are statistically significant with a large effect ($p < 0.001$; Glass' $\Delta > 0.8$). In terms of $M_{\text{Infections}}$, the *Ness* society has lower total number of infections (0.8909) than PRIMITIVE (48.3351), SANCTION (13.8398), HINT (2.2213), and TELL (20.4740) societies. The differences in the results are statistically significant with a large effect ($p < 0.001$; Glass' $\Delta > 0.8$).

With regard to $M_{\text{Vaccinated}}$, the *Ness* society has higher vaccination rate (98.7341) than PRIMITIVE (83.2589), SANCTION (36.743), HINT (11.1847) , and TELL (37.4301) societies. The differences in the results are statistically significant ($p < 0.001$; Glass' $\Delta > 0.8$). With $M_{\text{Vaccinated}}$, we observe that a vaccination norm emerges with a majority above 90% in HINT and *Ness* society. Specifically, even without a top-down imposed shared expectation on vaccination, most agents in HINT and *Ness* society learn that vaccination maximizes their payoff.

Table 3.6: Comparing the *Ness* society with baseline agent societies on various metrics and their statistical analysis with Glass' Δ and p-value. The metric row for each hypothesis shows the numeric value of the metric after simulation convergence. The performance in disease control among agent societies is based on attitudes expressed and information shared as in the order of *Ness*, HINT, TELL, SANCTION, and PRIMITIVE societies. However, the results of vaccination are in reverse order. HINT and *Ness* societies have similar isolation and forced quarantine rate, which is higher than TELL, SANCTION societies. Due to similar isolation and forced quarantine rate, The emotions measured between HINT and *Ness* societies are close. The goal satisfaction among agent societies is ranked in the order of *Ness*, HINT, TELL, SANCTION, and PRIMITIVE societies.

		PRIMITIVE	SANCTION	HINT	TELL	*Ness*
$H_{\text{Disease control}}$	M_{Infected}	12.6232	2.6343	0.1349	4.2050	0.1766
	p-value	< 0.001	< 0.001	0.4420	< 0.001	–
	Δ	−0.9703	−0.2688	0.0277	−0.3282	–
	M_{Healthy}	46.3426	77.6017	99.0701	65.0820	98.6826
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	–
	Δ	17.6160	3.4038	−0.2654	4.7757	–
	M_{Deceased}	41.0343	19.7640	0.7951	30.7130	1.1408
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	–
	Δ	−3.3430	−6.1093	7.0086	−5.3093	–
	$M_{\text{Infections}}$	48.3351	13.8398	0.7967	20.4740	1.0384
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	–
	Δ	−2.6566	−6.8475	13.5771	−5.7993	–
	$M_{\text{Vaccinated}}$	83.2589	36.7430	94.8081	37.4301	96.4632
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	–
	Δ	1.5930	17.5198	0.2543	12.7733	–
$H_{\text{Isolation}}$	$M_{\text{Isolation}}$	0.6162	0.9646	0.9978	0.9343	0.9975
	p-value	< 0.001	< 0.001	0.6667	< 0.001	–
	Δ	1.6388	0.3216	−0.0141	0.4472	–
	$M_{\text{Forced quarantine}}$	–	0.0258	0.0003	0.0401	0.0004
	p-value	–	< 0.001	0.4225	< 0.001	–
	Δ	–	−0.2657	0.0299	−0.3114	–
H_{Goal}	M_{Goal}	0.1875	0.2618	0.3118	0.2273	0.3070
	p-value	< 0.001	< 0.001	< 0.001	< 0.001	–
	Δ	3.0343	3.1834	−0.4581	3.7547	–

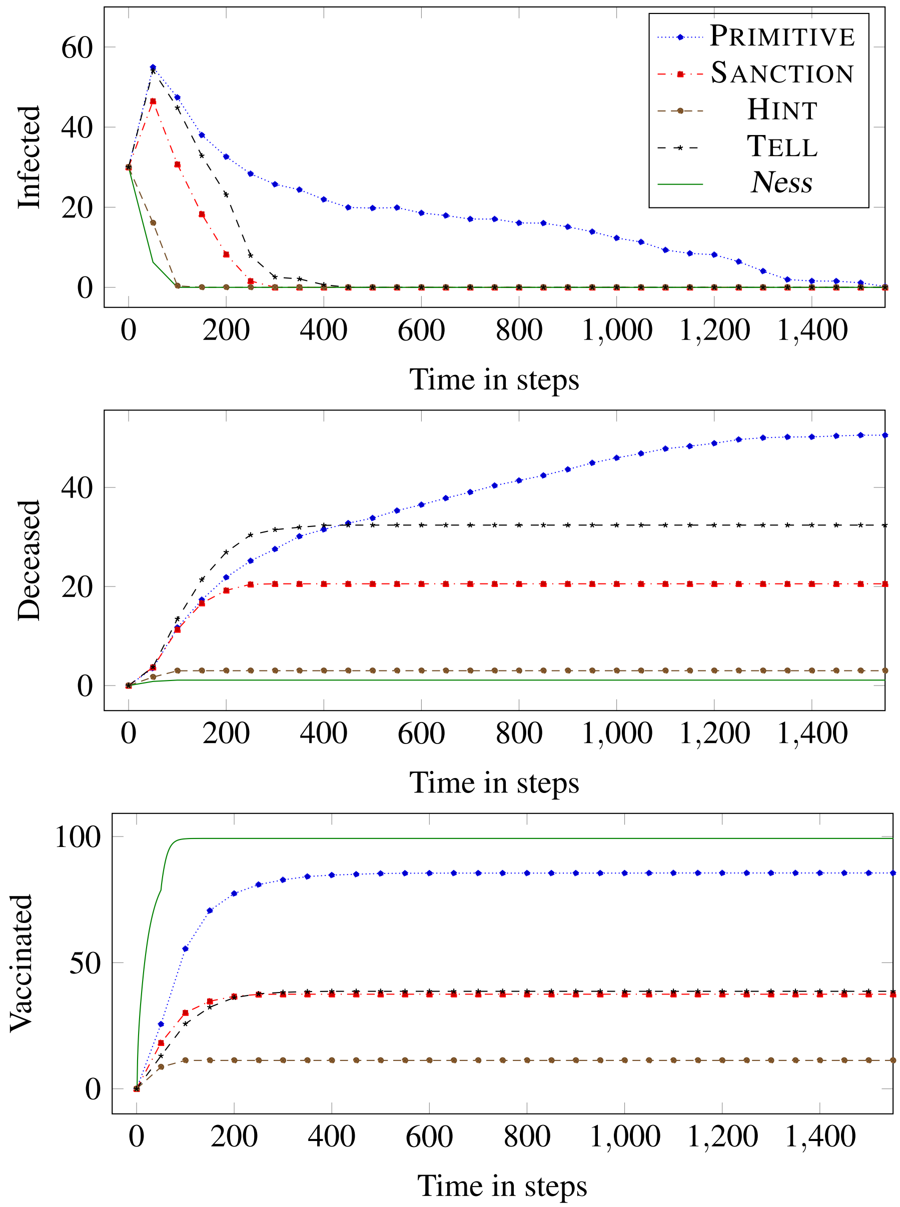


Figure 3.3: The *Ness* society has the least infected and deceased agents with the highest vaccination rate among all societies significantly. However, despite fewer vaccinated agents, the HINT agent society has significantly fewer infected agents, fewer deceased, and more healthy agents ($p < 0.001$; Glass' $\Delta > 0.8$) than the other baseline societies. The number of infected and deceased agents in *Ness* is lower than HINT and the number of healthy agents in *Ness* is higher than HINT but lines here overlap because of scaling. (Appendix 3.7 includes plots for the first 500 steps where the differences are noticeable.)

3.4.2 $H_{Isolation}$

To evaluate $H_{Isolation}$, we measure the proportion of isolation ($M_{Isolation}$) and the number of agents in forced quarantine ($M_{Forced\ quarantine}$) and the percentage of infected agents ($M_{infected}$) in agent societies. Whereas isolation means staying home when infectious, forced quarantine means being forced to stay home by others. Figure 3.4 exhibits plots comparing $M_{Isolation}$ and $M_{Forced\ quarantine}$. Note that we consider no infected agents remaining as all infected in isolation.

We observe that the *Ness* society has a higher tendency to stay isolated (0.9978) when infected than the PRIMITIVE (0.6162), SANCTION (0.9646), HINT (0.9934), and TELL (0.9343) societies. The differences in these results are statistically significant ($p < 0.001$; Glass' $\Delta >$ 0.8).

In terms of $M_{Forced\ quarantine}$, we notice that the *Ness* society has lower number of forced quarantine executed (0.0002) than SANCTION (0.0258), HINT (0.0008), and TELL (0.0401) societies. Whereas the results of statistical analysis shows that the differences between the *Ness* and HINT societies are statistically significant with a medium effect ($p < 0.01$; Glass' $\Delta > 0.5$), the differences between *Ness*, SANCTION, and TELL societies are significant ($p < 0.001$) but with small effect (Glass' $\Delta \approx 0.2$).

From $M_{Isolation}$ and $M_{Forced\ quarantine}$ and $M_{infected}$, we observe that a norm emerges with a majority above 90% in all societies other than PRIMITIVE society. Specifically, agents in societies with a prescriptive norm learn to comply with the norm and stay self-quarantined when infected. Furthermore, we see that this norm emerges fastest in *Ness* (0.9971) and HINT (0.9979) societies. With more subtle attitudes and information from emotions, agents in *Ness* and HINT societies learn faster than agents in TELL society.

While each society exhibits isolation behavior learning, this learning comes at different costs. Societies consider hints come at least costs, in terms of human resources and equipment costs from forced quarantine.

3.4.3 H_{Goal}

To evaluate H_{Goal}, we measure the goal satisfaction (M_{Goal}) in agent societies. Figure 3.5 shows the plot comparing M_{Goal} in agent societies. We observe that agents in the HINT society have the highest goal satisfaction (0.3213) than the PRIMITIVE (0.1875), SANCTION (0.2618), and TELL (0.2273) societies. The *Ness* society has slightly fewer goal satisfaction (0.3107) than the HINT society. The differences between M_{Goal} yielded by the *Ness* society and the PRIMITIVE, SANCTION, HINT, and TELL societies are statistically significant ($p < 0.001$; Glass' $\Delta > 0.8$).

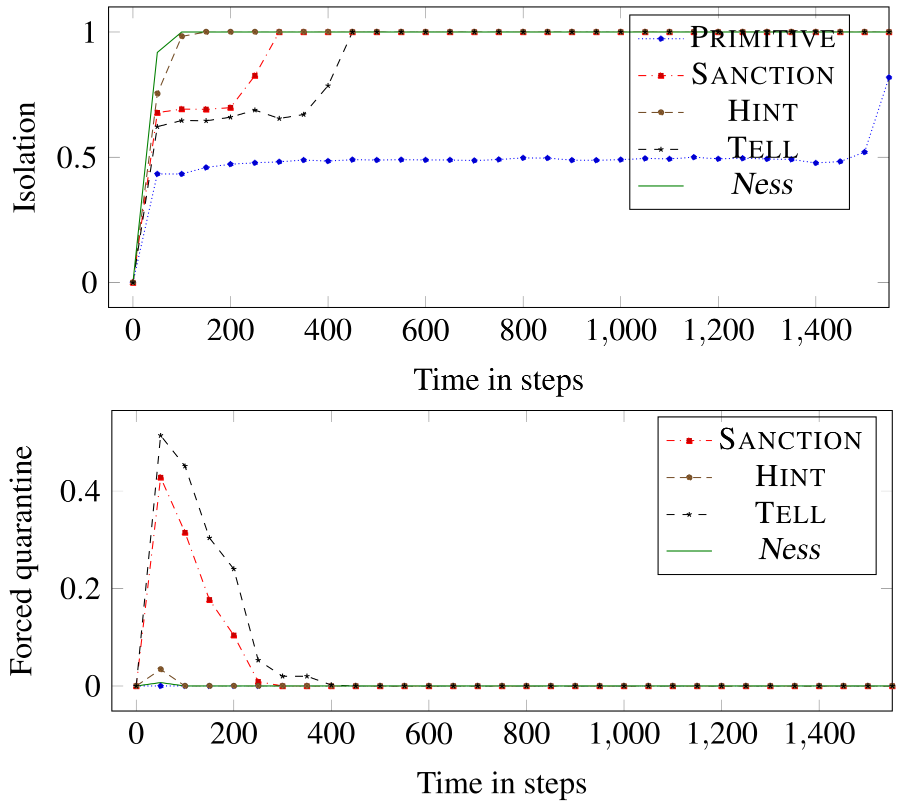


Figure 3.4: The percentage of isolation in the HINT and the *Ness* agent societies is significantly higher ($p < 0.001$; Glass' $\Delta > 0.8$) than the societies that do not consider hints. The *Ness* society has significantly fewer agents in forced quarantine ($p < 0.001$; Glass' $\Delta > 0.8$) to achieve stable cooperation than the SANCTION and the TELL societies. The difference between the *Ness* and HINT societies is not statistical significant and can be ignored ($p > 0.05$). The lines for *Ness* and HINT overlap.

3.5 Related Work

Research on norms and norm emergence closely relates to our contributions.

Andrighetto et al. (2013) show that a combination of verbal normative information, specifically positive normative content, and negative sanction leads to higher and more stable cooperation with human subjects and agent-based simulation. These models include normative reasoning but leave out soft signals such as hints. Kalia et al. (2019) demonstrate how signals such as emotions influence norm satisfaction. Hints in *Ness* could be understood as emotions but hints serve both as an sanctioning approach and providing information about norms.

Bourgais et al. (2019) present an agent architecture that integrates cognition, contagion, personality, norms, and social relations to simulate humans and ensure explainable behaviors.

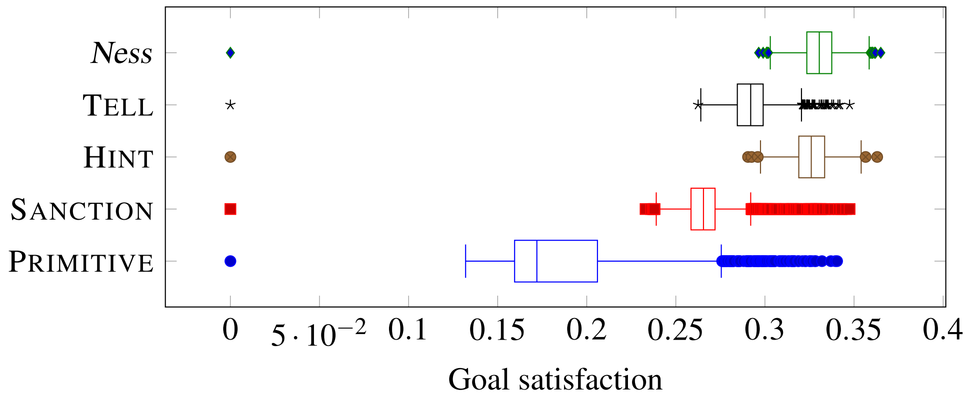


Figure 3.5: The *Ness* society yields significantly more goal satisfaction ($p < 0.001$; Glass' $\Delta > 0.8$) than the PRIMITIVE, SANCTION, and TELL societies. The differences in the results of the *Ness* and the HINT societies are small ($p < 0.001$; Glass' $\Delta < 0.2$).

Argente et al. (2022) propose an abstract normative emotional agent architecture, an extension of BDI architecture that combines emotional, normative, and cognitive component. Tzeng et al. (2021) combine normative model, a BDI model, and emotions for the decision-making process. Agents in *Ness* learn from their interactions with the environment and further interpret norms from various signals.

Dignum et al. (2020) associate the interventions that governments can take and their economic and social consequences with the SEIR model since effective and sustainable solutions cannot exist without considering these factors. *Ness* further takes social signaling into consideration.

de Mooij et al. (2022) develop a large-scale data-driven agent-based simulation model where each agent reasons about their internal attitudes and external factors to simulate behavioral interventions in the real world. *Ness* provides a framework that enables norm emergence and accommodates norms imposed by governments.

Dell'Anna et al. (2020) introduce a norm revision component that uses data collected from interactions and an estimation of agents' preferences to modify sanctions at runtime. de Lima et al. (2019) framework enables agents to pick sanctions appropriate to the context. Realpe-Gómez et al. (2018) present a model in which agents incorporate personal and normative considerations. Specifically, agents make decisions that maximize their respective payoffs while appraising their group's social norms. *Ness* defines the utility function based on normative information learned from social signaling.

Mukherjee et al. (2008) investigate the effects of heterogeneous agents using different learning algorithms. In addition, they study norm emergence when agent interactions are physically constrained. With *Ness* we study norm emergence when considering emotions from interactions. Airiau et al. (2014) model that supports the emergence of social norms by learning from interactions with a group of agents. In *Ness*, we include cognition via social signals in our agent model.

Hao et al. (2017) propose learning strategies based on local exploration and global exploration to support the emergence of social norms. Whereas their model focus on maximizing the average payoffs among agents, *Ness* focuses is on investigating influence of various signaling mechanisms.

Morales et al. (2018) focus on the stability of synthesized norms that are verified by an evolutionary process. Savarimuthu et al. (2010) propose an algorithm to identify obligation norms based on association rule mining, a data mining technique. Pernpeintner (2021) proposes a governance approach that restrict action spaces based on publicly observable behaviors and transitions.

Levy and Griffiths (2021) propose a framework that introduces congested actions where an agent's reward is not from pairwise interaction but is a function of others' actions and the environment. *Ness* enables social learning from personal observation or by normative information sharing from explicit messages or soft signals including hints.

3.6 Discussion

We present an approach that combines models of social signaling to address the emergence of norms. The novelty of our approach arises from its comprehensive treatment of the three main kinds of signals that drive norm emergence: sanctions, tell, and hint. Including normative information from messages or hints enables indirect social learning, which resembles human behaviors in the real world.

3.6.1 Summary of Findings

Our main findings are that agents who signal hints and respond to normative information converge to norms faster than those who respond only to hard sanctions or explicit communication of approval or disapproval. The agent societies that consider hints are also robust in complying to the converged norms compared to those who do not consider hints. Specifically, in our experiments, *Ness* and HINT societies exceed the 90% of norm emergence threshold faster than other societies and their compliance to the converged norm is higher than PRIMITIVE, SANCTION, and TELL societies.

Our pandemic environment simulation results show that (1) *Ness* society enables better control on the spread of disease than baseline societies, (2) agents in *Ness* and HINT societies learn the self-isolation norm faster and are more willing to isolate themselves when infected, (3) agents in *Ness* society have higher goal satisfaction than baseline societies. In summary, *Ness* agents effectively avoid undesirable results and yield higher satisfaction than baseline agents.

3.6.2 Limitations and Threats to Validity

We made simplifying assumptions that agents can infer each other's signals and that all signals are genuine and honest. These assumptions may not apply in all cases but are essential when interacting with other AI systems or needing special care. We would assume and expect AI systems, to be honest in human-robot interaction. In addition, the signals of people who need special care reflect their needs.

Our scenario involves human-human interaction, but other scenarios may consider purely artificial societies or diverse societies with humans and machines (as agents become capable enough to form and express emotions).

Although the consideration of norms is essential in human behaviors, humans can deviate from norms if such violation leads to better results. However, we focus on how social signals affect norm emergence by simplifying and assuming agents favoring norm-satisfaction behaviors.

3.6.3 Future Directions

While AI has been part of our daily lives nowadays, incorporating human ethics into AI becomes a necessary problem (Murukannaiah et al. 2020; Ajmeri et al. 2020; Lopez-Sanchez et al. 2017). Since human behavior is driven by the pursuit of values, studying human values helps us understand human decisions and create agents that reason over human values (Liscio et al. 2021). Social signals could also convey values. While Montes and Sierra (2021) automate norm synthesis based on value promotion, a interesting direction is to embed values into autonomous agents. That is, how can we develop agents that are capable of making value-aligned decisions. A line of future research is to investigate dimensions of emotions, physical arousal, that describes the strength of the emotional state. Another future direction includes considering a mix of personality types in *Ness*. We can investigate how different values influence human interactions in future research to support high heterogeneity.

3.7 Additional Results

Figure 3.6 plots the total number of infected agents in various agent societies.

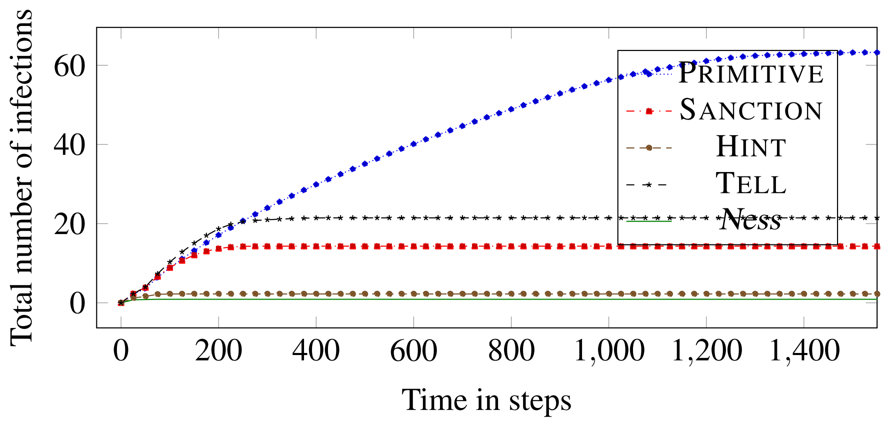


Figure 3.6: Comparing the total number of infections ($M_{\text{Infections}}$) in various agent societies. The *Ness* agent society has fewer total number of infections ($p < 0.001$; Glass' $\Delta > 0.8$) than the baseline societies.

Figures 3.7, 3.8, 3.9, and 3.10 shows plots for the number of infected, deceased, healthy, and vaccinated agents in the first 500 steps in various agent societies, where the differences are noticeable.

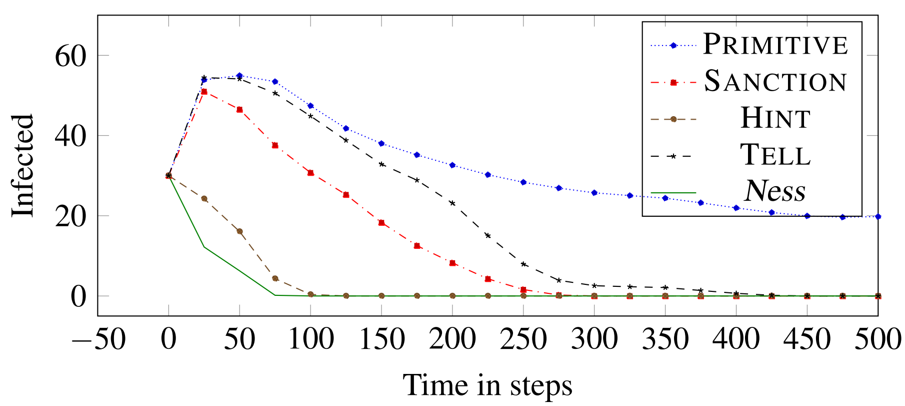


Figure 3.7: Comparing the number of infected agents in the first 500 steps. The differences are noticeable here.

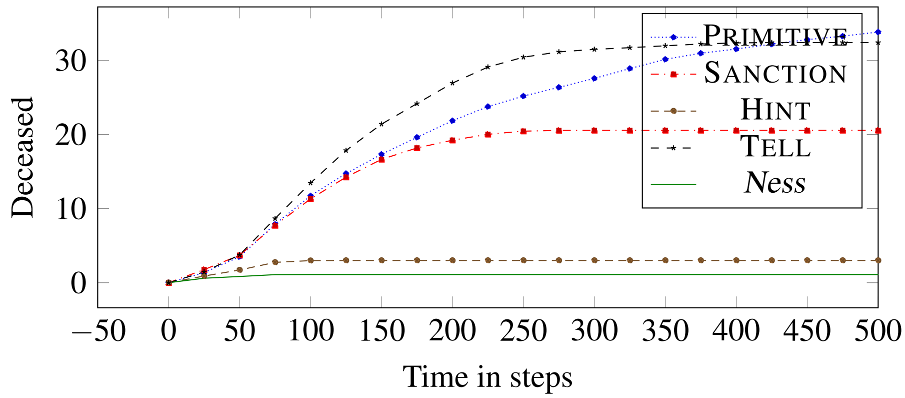


Figure 3.8: Comparing the number of deceased agents in the first 500 steps. The differences between *Ness* and HINT are noticeable here.

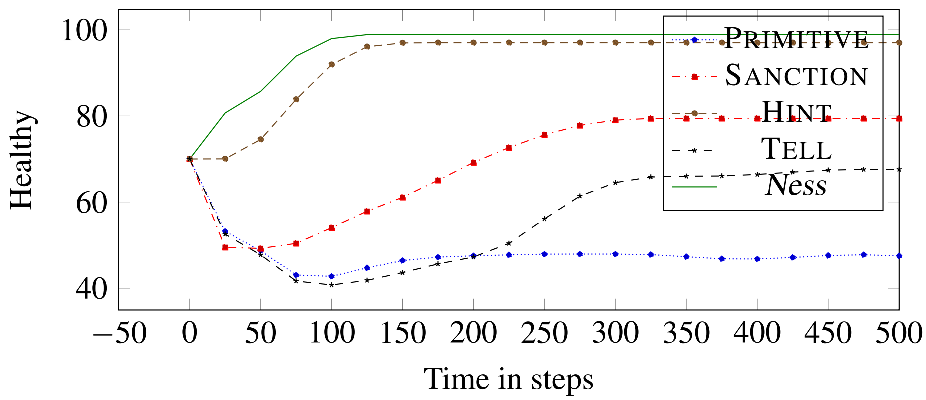


Figure 3.9: Comparing the number of healthy agents in the first 500 steps. The differences are noticeable here.

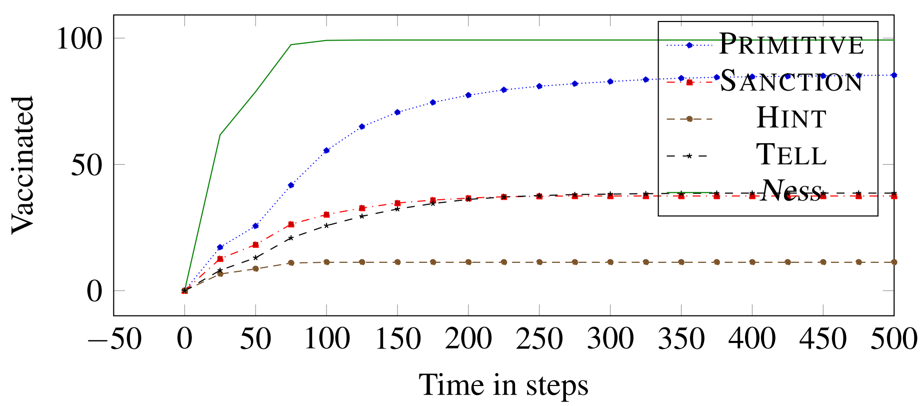


Figure 3.10: Comparing the number of vaccinated agents in the first 500 steps. The differences between *Ness* and other baseline agent societies are noticeable here.

Figure 3.11 shows the number of agents in self-isolation and forced-isolation in various agent societies in the first 500 steps of the simulation.

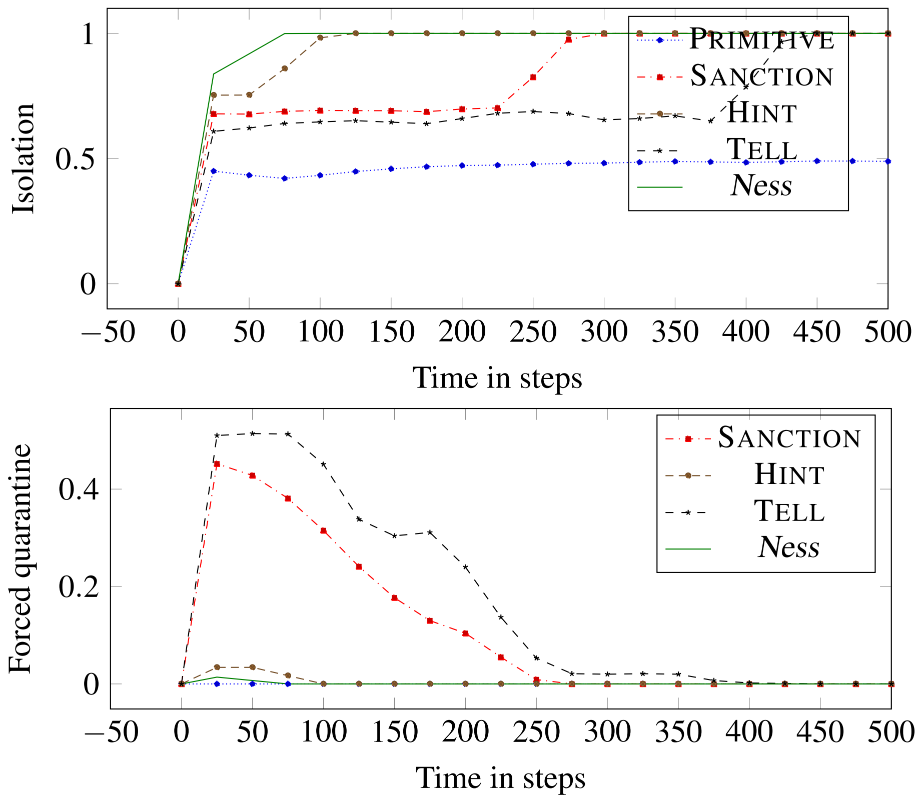


Figure 3.11: Comparing the number of agents in self-isolation ($M_{\text{Isolation}}$) and the number of agents in forced quarantine ($M_{\text{Forced quarantine}}$) in the *Ness* and baseline agent societies in the first 500 steps.

CHAPTER

4

SOCIAL VALUES ORIENTATION FOR ROBUST NORM EMERGENCE

4.1 Introduction

What makes people make different decisions? Schwartz (2012) defined ten fundamental human values, and each of them reflects specific motivations. Besides values, preferences define an individual's tendency to make a subjective selection among alternatives. Whereas values are relatively stable, preferences are sensitive to context and constructed when triggered (Slovic 1995).

In the real world, humans with varied weights of values evaluate the outcomes of their actions subjectively and act to maximize their utility (Schwartz 2012). In addition to values, an individual's social value orientation (SVO) influences the individual's behaviors (Van Lange 1999). Whereas values define the motivational bases of behaviors and attitudes of an individual (Schwartz 2012), social value orientation indicates an individual's preference for resource allocation between self and others (Griesinger and Livingston Jr. 1973). Specifically, social value orientation provides stable subjective weights for making decisions (Murphy and Ackermann

2014). When interacting with others is inevitable, one individual's behavior may affect another. SVO revises an individual's utility function by assigning different weights to itself and others. Here is an example of a real-world case of SVO.

Example 4 ***SVO.***

During a pandemic, the authorities announce a mask-wearing regulation and claim that regulation would help avoid infecting others or being infected. Although Felix tests positive on the pandemic and prefers not to wear a mask, he also cares about others' health. If he stays in a room with another healthy person, Elliot, Felix will put the mask on.

An agent is an autonomous, adaptive, and goal-driven entity (Russell and Norvig 2010). Whereas many works assume agents consider the payoff of themselves, humans may further consider social preferences in the real world. e.g., payoffs of others or social welfare (Charness and Rabin 2002). When humans are in the loop along with software, there are emerging need to consider human factors when building modern software and systems. These systems should consider human values and be capable of reasoning over humans' behaviors to be realistic and trustworthy.

In a multiagent system, social norms or social expectations (Rummel 1975; Ajmeri et al. 2017) are societal principles that regulate our behavior towards one another by measuring our perceived psychological distance. Humans evaluate social norms based on human values. Most previous works related to norms do not consider human values and assume regimented environments. However, humans are capable of deliberately adhering to or violating norms. Previous works on normative agents consider human values and theories on sociality (Ajmeri et al. 2020; Verhagen 2000) in decision-making process. SVO as an agent's preference in a social context has not been fully explored.

Contributions

We investigate the following research question.

RQ_{SVO}. How do the preferences for others' rewards influence norm compliance?

To address RQ_{SVO}, we develop *Fleur*, an agent framework that considers values, personal preferences, and social norms when making decisions. Our proposed framework *Fleur* combines world model, cognitive model, emotion model, and social model. Since values are abstract and need further definition, we start with social value orientations, the stable preferences for resource allocation, in this work. Specifically, *Fleur* agents take into account social value orientation in utility calculation.

Findings

We evaluate *Fleur* via an agent simulation of a pandemic scenario designed as an iterated single-shot and intertemporal social dilemma game. We measure compliance, social experiences, and invalidation during the simulation. We find that the understanding of SVO helps agents to make more ethical decisions.

Organization

Section 4.2 presents the related works. Section 4.3 describes the schematics of *Fleur*. Section 4.4 details the simulation experiments we conduct and the results. Section 4.5 presents our conclusion and directions for future extensions.

4.2 Related Works

Griesinger and Livingston Jr. (1973) present a geometric model of SVO, the social value orientation ring as Figure 4.2. Van Lange (1999) proposes a model and interprets prosocial orientation as enhancing both joint outcomes and equality in the outcomes. Declerck and Bogaert (2008) describe social value orientation as a personality trait. Their work indicates that prosocial orientation positively correlates with adopting others' viewpoints and the ability to infer others' mental states. On the contrary, an individualistic orientation shows a negative correlation with these social skills. *Fleur* follows the concepts of social preferences from (Griesinger and Livingston Jr. 1973).

Szekely et al. (2021) show that high risk promotes robust norms, which have high resistance to risk change. de Mooij et al. (2022) build a large-scale data-driven agent-based simulation model to simulate behavioral interventions among humans. Each agent reasons over their internal attitudes and external factors in this work. Ajmeri et al. (2018) show that robust norms emerge among interactions where deviating agents reveal their contexts. This work enables agents to empathize with other agents' dilemmas by revealing contexts. Instead of sharing contexts, values, or preferences, *Fleur* approximates others' payoff with observation. Serramia et al. (2018) consider shared values in a society with norms and focus on making ethical decisions that promote the values. Ajmeri et al. (2020) propose an agent framework that enables agents to aggregate the value preferences of stakeholders and make ethical decisions accordingly. This work takes other agents' values into account when making decisions. Mosca and Such (2021) describe an agent framework that aggregates the shared preferences and moral values of multiple users and makes the optimal decisions for all users. Kalia et al. (2019) investigate the relationship between norm outcomes and trust and emotions. Tzeng et al. (2021) consider emotions as sanctions. Specifically, norm satisfaction or norm violation may trigger self-directed and other-directed emotions, which further enforce social norms. Dell'Anna et al. (2019) propose a mechanism to regulate a multiagent system by revising the sanctions at runtime to achieve runtime norm enforcement.

Agrawal et al. (2022) provide and evaluate explicit norms and explanations. Winikoff et al. (2021) construct comprehensible explanations with beliefs, desires, and values. Kurtan and Yolum (2021) estimate privacy values with existing shared images in a user's social network. Tielman et al. (2019) derive norms based on values and contexts. However, these works do not consider the differences between agents and the influences of an individual's behavior on others. Mashayekhi et al. (2022) model guilt based on inequity aversion theory for an individual perspective on prosociality. In addition, they consider justice from a societal perspective on prosociality. Whereas Mashayekhi et al. (2022) assume agents may be self-interested and their decisions may be affected by others' performance, *Fleur* investigates the influence of social value orientations.

Table 4.1 summarizes related works on ethical agents. Adaptivity describes the capability of responding to different contexts. Empathy defines the ability to consider others' gain. The information share indicates information sharing among agents. The information model describes the applied models to process information and states. Among varied information models, contexts describe the situation in which an agent stands. Emotions are the responses to internal or external events or objects. Guilt is an aversive self-directed emotion. Explicit norms state causal normative information, including antecedents and consequences. Values and preferences both define desirable or undesirable states.

4.3 *Fleur*

We now discuss the schematics of *Fleur* agents.

Figure 4.1 shows the architecture of *Fleur*. *Fleur* agents consists of five main components: cognitive model, emotion model, world model, social model, and a decision module.

4.3.1 Cognitive Model

Cognition relates to conscious intellectual activities, such as thinking, reasoning, or remembering, among which human values and preferences are essential. Specifically, values and preferences may change how an individual evaluates an agent, an event, or an object. In *Fleur*,

Table 4.1: Comparisons of works on ethical agents with norms and values.

Research	Adaptivity	Empathy	Information Share	Information Model
Fleur	✓	✓	✗	Preferences & Emotions & Contexts
Agrawal et al. (2022)	✓	✗	✓	Explicit norms
Ajmeri et al. (2018)	✓	✓	✓	Contexts
Ajmeri et al. (2020)	✓	✓	✓	Values & Value preference & Contexts
Kalia et al. (2019)	✓	✗	✗	Trust & Emotions
Kurtan and Yolum (2021)	✓	✗	✗	Values
Mashayekhi et al. (2022)	✓	✓	✓	Guilt
Mosca and Such (2021)	✓	✓	✓	Preferences & Values
Serramia et al. (2018)	✓	✗	✗	Values
Tielman et al. (2019)	✓	✗	✓	Values & Contexts
Tzeng et al. (2021)	✗	✗	✗	Emotions
Winikoff et al. (2021)	✓	✗	✗	Values & Beliefs & Goals

We start with including human preferences. While preferences are the attitudes toward a set of objects in psychology (Slovic 1995), individual and social preferences provide intrinsic rewards. For instance, SVO provides agents with different preferences over resource allocations between themselves and others. Figure 4.2 demonstrates the reward distribution of different SVO types. The horizontal axis measures the resources allocated to oneself, and the vertical axis measures the resources allocated to others. Let $\overrightarrow{R} = (r_1, r_2, \ldots, r_n)$ represent the reward vector for a group of agents with size n. The reward for agent i considering social aspect is:

$$reward_i = r_i \cdot \cos\theta + r_{-i} \cdot \sin\theta \tag{4.1}$$

where r_i represents the reward for agent i and r_{-i} is the mean reward of all other agents interacting with agent i. Here we adopt the reward angle in (McKee et al. 2020) and represent agents' social value orientation with θ. We define $\theta \in \{90°, 45°, 0°, -45°\}$ as SVO $\in$ {altruistic, prosocial, individualistic, competitive}, respectively. With the weights provided by SVO, the presented equation enables the accommodation of social preferences.

In utility calculation, we consider two components: (1) extrinsic reward and (2) intrinsic reward. Whereas extrinsic rewards come from the environment, intrinsic rewards stem from

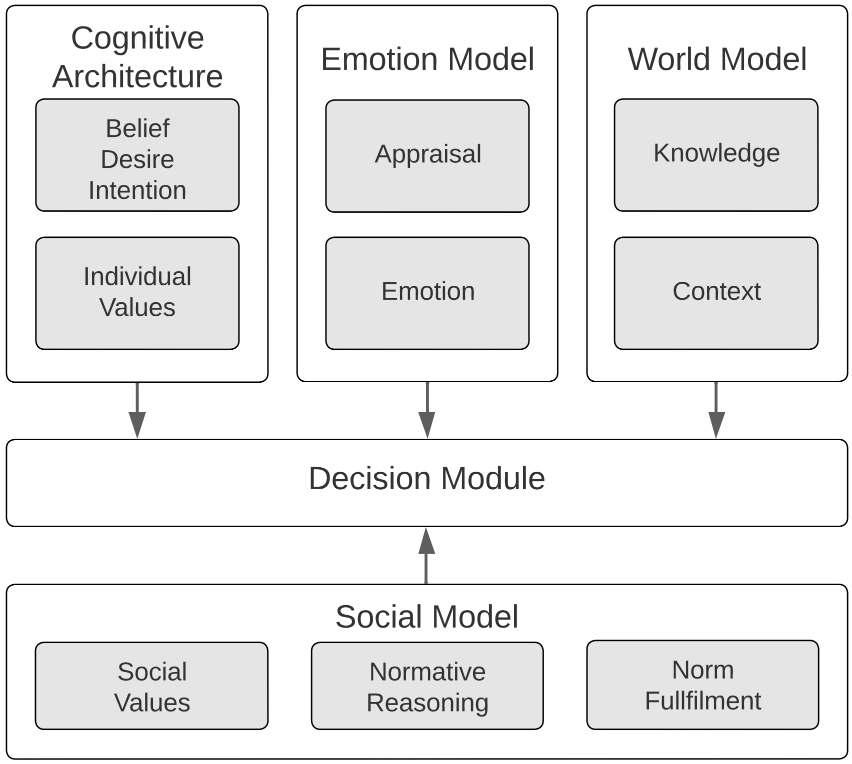


Figure 4.1: *Fleur* architecture.

internal stats, e.g., human values and preferences.

We extend the Belief-Desire-Intention (BDI) architecture (Rao and Georgeff 1991). An agent forms beliefs based on the information from the environment. The desire of an agent represents having dispositions to act. An agent's intention is a plan or action to achieve a selected desire.

Take Example 4 for instance. Since Felix has an intention to maximize the joint gain with Elliot, he may choose a strategy to not increase his payoff at the cost of others' sacrifice.

4.3.2 Emotion Model

We adopt the OCC model of emotions (Ortony et al. 1988). Specifically, our emotion model appraises an object, an action, or an event and then triggers emotions. We consider emotional valence and assume norm satisfaction or norm violation yields positive or negative emotions if self behaviors align with the norms.

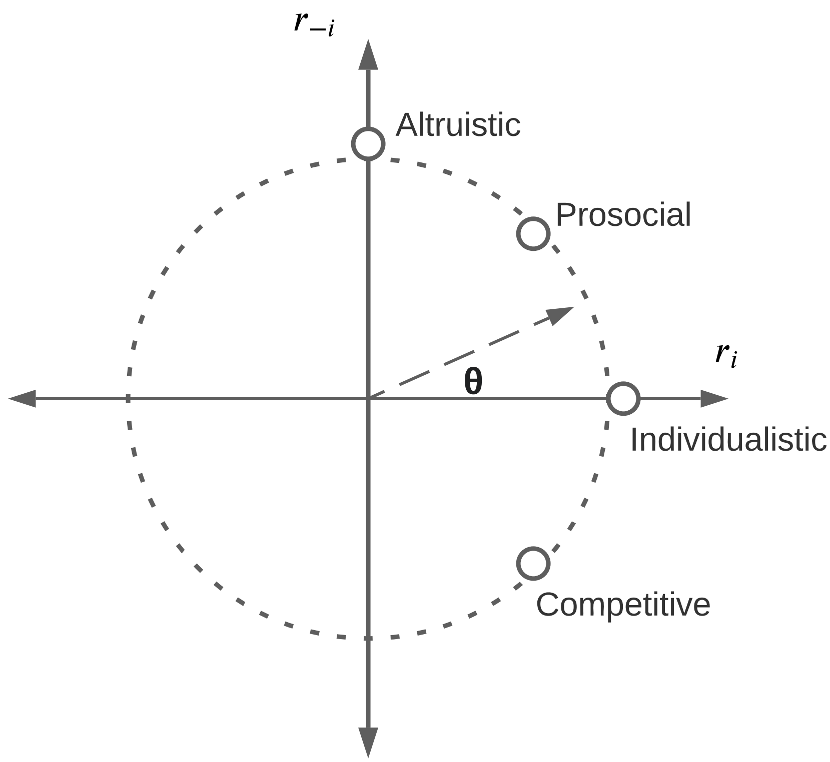


Figure 4.2: Representation of Social Value Orientation (Griesinger and Livingston Jr. 1973; McKee et al. 2020). r_i denotes outcome for oneself and r_{-i} denotes outcomes for others.

4.3.3 World Model

The world model describes the contexts in which *Fleur* agents stand and represents the general knowledge *Fleur* agents possess. A context is a scenario that an agent faces. Knowledge in this model are facts of the world. In Example 4, the context is that an infected individual, Felix, seeks to maximize the collective gain of himself and a healthy individual, Elliot. In the meantime, Felix acknowledges that a pandemic is ongoing.

4.3.4 Social Model

The social model of an agent includes social values, normative reasoning, and norm fulfillment. Social values define standards that individuals and groups employ to shape the form of social order (Tsirogianni et al. 2014), e.g., fairness and justice. Agents use the normative-reasoning component to reason over states, norms, and possible outcomes of satisfying or violating norms. Norm fulfillment checks if a norm has been fulfilled or violated with the selected action. Sanctions may come after norm fulfillments or violations.

4.3.5 Decision Module

The decision module selects actions based on agents' payoffs and individual values. We apply Q-Learning (Watkins and Dayan 1992), a model-free reinforcement learning algorithm that learns from trial and error, to our agents. Q-Learning approximates the action-state value $Q(s,a)$ (Q value), with each state and action:

$$Q'(s_t,a_t) = Q(s_t,a_t) + \alpha * (R_t + \gamma \max_{a'} Q(s_{t+1},a) - Q(s_t,a_t)) \tag{4.2}$$

where $Q'(s_t,a_t)$ represents the updated Q-value after performing action a at time t and s_{t+1} represents the next state. α denotes the learning rate in the Q-value update function, and R_t represents the rewards received at time t after acting a. γ defines the reward discount rate, which characterizes the importance of future rewards. Agents observe the environment, form their beliefs about the world, and update their state-value with rewards via interactions. By approximating the action-state value, the Q-Learning algorithm finds the optimal policy via the expected and cumulative rewards.

Algorithm 3 describes the agent interaction in our simulation.

4.4 Experiments

We now describe our experiments and discuss the results.

4.4.1 Experimental Scenario: Pandemic Mask Regulation

We build a pandemic scenario as an iterated single-shot and intertemporal social dilemma. We assume that the authorities have announced a masking regulation. In each game, each agent selects from the following two actions: (1) wear a mask, and (2) not wear a mask. Each agent has its inherent preferences and social value orientation. An agent forms a belief about its partner's health based on its observation. During the interaction, the decision an agent makes affects itself and others. The collective behaviors among agents determine the the dynamics in a society. Each agent receives the final points from its own action and effects from others:

Algorithm 3: Decision loop of a *Fleur* agent

Initialize one agent with its desires D and preference P and SVO angle θ;
Initialize action-value function Q with random weights w;
for t=*1,T* **do**
    Pair up with another agent pn to interact with;
    Observe the environment (including the partner and its θ) and form beliefs b_t;
    With a probability ε select a random action a_t
      Otherwise select $a_t = argmax_a Q(b_t, a; w)$
    Execute action a_t and observe reward r_t;
    Observe the environment (including the partner) and form beliefs b_{t+1};
    Activate norms N with beliefs b_t, b_{t+1}, and action a_t;
    if $N \,! = \varnothing$ **then**
        Sanction the partner based on a_t and its behavior;
    end
end

$R_{sum} = P_{i\_self} + P_{i\_other} + S_j$. P_{i_self} denotes the payoff from the action that agent i selects considering the reward distribution in Figure 4.2 and self-directed emotions. P_{i_other} is the payoff from the action that the other agent performs. S_j denotes the other-directed emotions from others towards agent i.

Table 4.2: Payoff for an actor and its partner based on how the actor acts and how its action influence others. Column Actors show the points from the actions of the actor. Column Partners display the points from the actions to the partner.

Health		Actions			
		Mask		No mask	
Actor	Partner	Actor	Partner	Actor	Partner
healthy	healthy	0.00	0.00	0.00	0.00
healthy	infected	1.00	0.00	−1.00	0.00
infected	healthy	0.00	1.00	0.00	−1.00
infected	infected	0.50	0.50	−0.50	−0.50

Table 4.3: Payoff for decisions on preferences.

Type	Decisions	
	Satisfy	Dissatisfy
Preference	0.50	0.00

Table 4.4: Payoff for decisions on norms.

Actor	Partner	
	Wear	Not-Wear
Wear	0.10	-0.10
Not-Wear	0.00	0.10

4.4.2 Experimental Setup

We develop a simulation using Mesa (Masad and Kazil 2015), an agent-based modeling framework in Python for creating, visualizing, and analyzing agent-based models. We ran the simulations on a device with 32 GB RAM and GPU NVIDIA GTX 1070 Ti.

We evaluated *Fleur* via a simulated pandemic scenario where agents' behaviors influence the collective outcome of the social game. A game-theoretical setting may be ideal for validating the social dilemma with SVO and norms. However, real-world cases are usually non-zero-sum games where one's gain does not always lead to others' loss. In our scenario, depending on the context, the same action may lead to different consequences for the agent itself and its partner. For instance, when an agent is healthy and its partner is infected, wearing a mask gives the agent a positive payoff from the protection of the mask but no payoff for its partner. Conversely, not wearing a mask leads to a negative payoff for the agent and no payoff for its partner. The payoff given to the agent and its partner corresponds to the X and Y axis in Figure 4.2. When formalizing social interactions with SVO in game-theoretical settings, the payoffs of actions for an agent and others are required information.

We incorporated beliefs and desires, and intentions into our agents. An agent observes its environment and processes its perception, and forms its beliefs about the world. In each episode, agents pair up to interact with one another and sanction based on their and partners' decisions (Table 4.4).

Context. A context is composed of attributes from an agent and others and the environment as shown in Table 4.2. We frame the simulation as a non-zero-sum game where one's gain does not necessarily lead to the other parties' loss.

Preference. In psychology, preferences refer to an agent's attitudes towards a set of objects. In our simulation, we set 40% of agents to prefer to wear and prefer not to masks individually. The rest of the agents have a neutral attitude on masks. The payoffs for following the preferences are listed in Table 4.3.

Social Value Orientation. Social value orientation defines an agent's preference for allocating resources between itself and others. We consider altruistic, prosocial, individualistic, and competitive orientations selected from Figure 4.2.

4.4.3 Hypotheses and Metrics

We compute the following measures to address our research question $\mathrm{RQ_{SVO}}$.

Compliance The percentage of agents who satisfy norms

Social Experience The total payoff of the agents in a society

Invalidation The percentage of agents who do not meet their preferences in a society

To answer our research question $\mathrm{RQ_{SVO}}$, we evaluate three hypotheses that correspond to the specific metric, respectively.

$\mathbf{H_{Compliance}}$**:** Preferences for others' rewards positively affect norm compliance with prosocial norms

$\mathbf{H_{Social\ Experience}}$**:** The distribution of preferences for others' rewards positively affect social experiences in a society

$\mathbf{H_{Invalidation}}$**:** Preferences for others' rewards negatively affect the tendency to meet personal preferences

4.4.4 Experiments

We ran a population of $N = 40$ agents in which we equally distributed our targeted SVO types: altruistic, prosocial, individualistic, and competitive. Since each game is a single-shot social dilemma, we consider each game as an episode. The training last for 500,000 episodes. In evaluation, we run 100 episodes and compute the mean values to minimize deviation from coincidence. We define our five societies as below.

Mixed society A society of agents with mixed social value orientation distribution

Altruistic society A society of agents who make decisions based on altruistic concerns

Prosocial society A society of agents who make decisions based on prosocial concerns

Selfish society A society of agents who make decisions based on selfish concerns

Competitive society A society of agents who make decisions based on competitive concerns

We assume all agents are aware of a mask-wearing norm. Agents who satisfy the norm receive positive emotions from themselves and others, as in Table 4.4. Conversely, norm violators receive negative emotions. Table 4.5 summarizes results of our simulation.

Table 4.5: Comparing agent societies with different social value orientation distribution on various metrics and their statistical analysis with Glass' Δ and p-value. Each metric row shows the numeric value of the metric after simulation convergence.

		Compliance	Social Experience	Invalidation
S_{mixed}	Results	63.40%	0.4483	0.2960
	p-value	–	–	–
	Δ	–	–	–
$S_{altruistic}$	Results	69.70%	0.5543	**0.3340**
	p-value	< 0.001	< 0.001	< 0.001
	Δ	0.6602	0.6116	0.4635
$S_{prosocial}$	Results	**70.25%**	**0.5656**	0.3228
	p-value	< 0.001	< 0.001	< 0.05
	Δ	0.7178	0.6771	0.3263
$S_{selfish}$	Results	65.10%	0.4695	0.2690
	p-value	0.2180	0.4245	< 0.05
	Δ	0.1781	0.1221	0.3293
$S_{competitive}$	Results	54.08%	0.2208	0.2888
	p-value	< 0.001	< 0.001	0.5412
	Δ	0.9772	1.3131	0.0884

Figure 4.3 displays the compliance, the percentage of agents who satisfy norms, in the mixed and baseline-agent societies. We find that the compliance in the altruistic and prosocial-agent society, averaging at 69.70% and 70.25%, is higher than in the mixed (63.34%) and

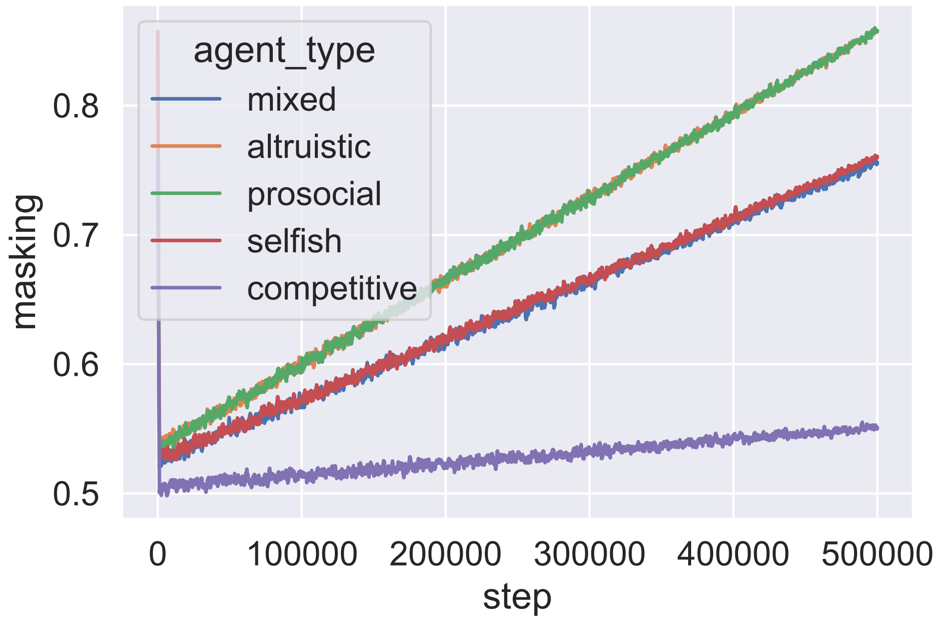


Figure 4.3: Compliance in training phase: The percentage of norm satisfaction in a society.

agent societies have no positive weights on others' payoff (65.10% and 54.08% for selfish and competitive-agent societies, respectively). The differences in the results of altruistic and prosocial-agent societies are statistically significant with medium effect ($p < 0.001$; Glass' $\Delta > 0.5$). Conversely, the competitive-agent society has the least compliance, averaging at 54.08%, with $p < 0.001$ and Glass' $\Delta > 0.8$. The results of the selfish-agent society (65.10%) shows no significant difference with $p > 0.05$ and Glass' $\Delta \approx 0.2$.

There are 25% of agents in the mixed-agent society are competitive agents. Specifically, they prefer to minimize others' payoff. A competitive infected agent may choose not to wear a mask when interacting with other healthy agents in this scenario. In the meantime, the selfish agents would maximize their self utility without considering others. Therefore, the behaviors of selfish and competitive agents may decrease compliance in the mixed-agent society.

Figure 4.4 compares the average payoff in the mixed and baseline-agent societies. The social experience in the altruistic and prosocial-agent society, averaging at 0.5543 and 0.5656, is higher than in the mixed (0.4483) and agent societies have no positive weights on others' payoff (46.95% and 22.08% for selfish and competitive-agent societies, respectively). The differences in the results of altruistic and prosocial-agent societies are statistically significant with medium effect ($p < 0.001$; Glass' $\Delta > 0.5$). On the contrary, the competitive-agent society has the least social experience, averaging at 0.2208, with $p < 0.001$ and Glass' $\Delta > 0.8$. The results of the

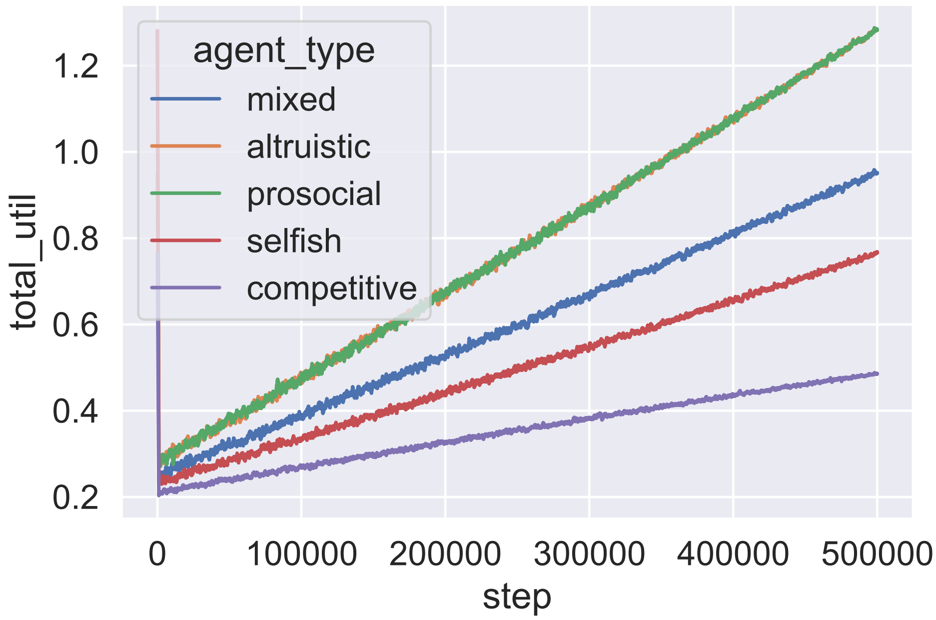


Figure 4.4: Social Experience in training phase: The total payoff of the agents in a society.

selfish-agent society (0.4695) shows no significant difference with $p > 0.05$ and Glass' $\Delta < 0.2$.

The mixed-agent society shows similar results as the selfish-agent society. Although 50% of the mixed-agent society agents are altruistic and prosocial, the competitive agents would choose to minimize others' payoff without hurting their self-interests.

Since the selfish agents do not care about others, they would act for the sake of their benefit. The selfish and competitive behaviors diminish the social experiences in society.

Figure 4.5 compares invalidation, the percentage of agents who do not meet their preferences in the mixed and baseline-agent societies.

The invalidation in the altruistic and prosocial-agent society, averaging at 33.40% and 32.28%, is higher than in the mixed (29.60%) and agent societies have no positive weights on others' payoff (26.90% and 28.88% for selfish and competitive-agent societies, respectively). The differences in the results of altruistic and prosocial-agent societies are statistically significant with small or medium effect ($p < 0.001$; Glass' $\Delta > 0.2$). On the contrary, the selfish-agent society has the least invalidation, average at 26.90%, with $p < 0.05$ and Glass' $\Delta > 0.2$. The results of the competitive-agent society (28.88%) shows no significant difference with $p > 0.05$ and Glass' $\Delta < 0.2$.

While agents who consider others' rewards positively achieve better compliance and social experiences, these achievements are based on their sacrifice of preferences. The altruistic and

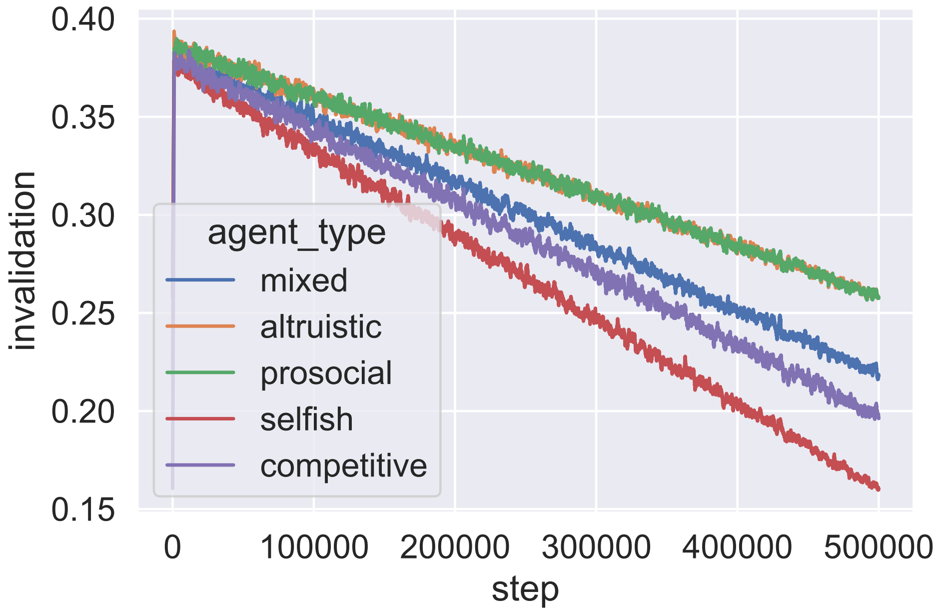


Figure 4.5: Invalidation in training phase: The percentage of agents who do not meet their preferences in a society.

prosocial agent societies have the most percentage of agents who do not meet their preferences.

4.4.5 Threats to Validity

First, our simulation has a limited action space. Moreover, different actions may have the same payoff in some contexts. Other behaviors may better describe different types of SVO, yet our focus is on showing how SVO influences normative decisions.

Second, we represent actual societies as simulations. While differences in preference and SVO among people are inevitable, we focus on validating the influence of SVO.

Third, to simplify the simulation, we assume fixed interaction, whereas real-world interactions tend to be random. An agent may interact with one another in the same place many times or have no interaction. We randomly pair up all agents to mitigate this threat and average out the results.

4.5 Conclusions and Directions

We present an agent architecture that integrates cognitive architecture, world model, and social model to investigate how social value orientation influences compliance with norms. We simulate a pandemic scenario in which agents make decisions based on their individual and social preferences. The simulations show that altruistic and prosocial-agent societies comply better with the mask norm and bring out higher social experiences. However, altruistic and prosocial agents trade their personal preferences for compliance and social experiences. The results between the mixed and selfish-agent societies show no considerable difference. The competitive agents in the mixed-agent society may take the responsibility.

Future Directions

Our possible extensions include investigating an unequal distribution of SVO in *Fleur* and applying real-world data in the simulation. Other future directions are incorporating values into agents, and revealing adequate information to explain and convince others of inevitable normative deviations (Agrawal et al. 2022; Murukannaiah et al. 2020; Woodgate and Ajmeri 2022).

CHAPTER

5

DECISION AND RATIONALE WITH VALUES

5.1 Introduction

When humans are in the loop, there are some challenges AI systems have to tackle (Woodgate and Ajmeri 2022). First, AI must adapt to the dynamic multiagent system (MAS), where agents come and go. Social norm defines the shared understanding of acceptable behaviors, which govern members' behaviors in a MAS. The norms governing the MAS must evolve to adapt to the changing environment and agents. Whereas other research focuses on synthesizing norms at runtime (Morris-Martin et al. 2023), we focus on the adaptability of agents. Second, there is an emergent need for AI systems to reason about human factors such as preferences and values. To be interpretable and trustworthy to humans, AI systems should make decisions aligned with human values (Liscio et al. 2023) at the micro level. Whereas the macro level of values focuses on the governance of MAS, the micro level of values focuses on the behaviors of individual agents. In decision-making, values define the motivational bases of an individual's behaviors and attitudes (Schwartz 2012), leading to different decisions and goals between agents. An agent who cares about free will may derive a goal of following its personal preferences from its value preferences. Whereas values are context-specific, values may transcend contexts. While contexts may vary, values provide a consistent framework to evaluate the decisions. For instance, an individual who has honesty as one of the core values will endeavor to be honest in various contexts, e.g., personal relationships, school, and work. In addition, different individuals may have different values. Although most people value personal health during a pandemic, some claim their freedom by not wearing masks or vaccinating.

To be trusted by humans, an agent should be able to provide rationales for their decisions (Winikoff et al. 2021; Ayci et al. 2023). Although social norms define social expectations and coordinate agent behaviors in a multiagent system, an agent may legitimately deviate from a norm (Singh and Singh 2023). Violating norms may lead to negative sanctions. At the macro level, values can yield norms to govern multiagent systems (Noriega et al. 2021; Liscio et al. 2023). Whereas individual agents act based on partial observations, justifying behavior via revealing information may enhance individual gain and resolve social conflict (Ajmeri et al. 2018).

Example 5 ***Reveal information.*** *Alice comes to office with a mask where she notices Bella not wearing a mask. Bella justifies her decision by stating that, first, there is no mask mandate in the office as the surrounding environment is safe. Second, she hates wearing a mask because wearing one causes her eczema. Alice agrees with Bella's perspective and does not avoid Bella.*

Langley (2019) and Miller (2019) suggest that the goal of rationales is to provide necessary information for a decision. In practice, information sharing provides extra information from an individual that others may be unable to observe, e.g., beliefs and preferences. However, what might it be like to live in a world where personal information and activity history are accessible everywhere? People who care about privacy might feel invaded or uncomfortable in that world. In addition, verbose rationales may be diverging and not convincing, leading to information overload. Therefore, here comes the question: What would be appropriate information sharing without sacrificing personal values? Besides personal feelings, information revealing can be risky because one's opponent can exploit the revealer's vulnerability in a competitive context.

Example 6 ***Adaptive information.***

Bella and Alice are enthusiastic about health. Although Bella prefers not to wear a mask due to possible skin issues, revealing that information may not be necessary. Bella justifies her behavior of not wearing a mask by stating that the surrounding environment is safe and that a mask is not needed. Alice finds Bella's rationale convincing and does not avoid Bella.

When humans are involved in the MAS, a good decision made by AI systems should go beyond physical gain and be aligned with human values. Whereas values guide motivations and drive decisions, rationales or information aligned with values best justify one's behaviors. In addition, values reflect different concerns in decision-making and conflict resolution among agents. Whereas value-aligned rationales enables agents to justify their behavior adaptively, deliberating over others' values may increase convincingness and acceptance. Based on the preceding intuition, we investigate the following research questions.

$\textbf{RQ}_{\textbf{Goal Adherence}}$. Do value-aligned rationales increase adherence to the original goal?

$\textbf{RQ}_{\textbf{Conflict Resolution}}$. Do value-aligned rationales increase the social resolution?

$\textbf{RQ}_{\textbf{Privacy Loss}}$. Does value-aligned rationales reduce privacy loss?

We address these questions and make the following contributions.

Contributions

First, we propose *Exanna*, a framework that incorporates values in decision-making, rationale generation, and reasoning over rationale. Second, whereas other works focus on making agent decisions interpretable to humans, *Exanna* agents provide rationales to both agents and humans.

Findings

We evaluate *Exanna* with simulations of a pandemic scenario. We consider agent societies with three characteristics for producing rationales: share all, share decision rules, and share value-aligned rules. With *Exanna*, we find that agents who consider value preferences when giving rationals deviate from goals and achieve higher conflict resolution. Surprisingly, rationals with less private information lead to better social experiences.

Novelty

This work presents essential perspectives on decisions and rationales with values. First, an individual's decision-making and evaluation involves values (Schwartz 2012), but values with higher weights dominate the decision. In addition, a compelling rationale often states causal relationships with the esteemed values of agents. An *Exanna* agent makes decisions based on its or its stakeholder's values. Upon generating rationales, the agent discloses the causal effects based on values, which are interpretable to both agents and humans. Second, individuals evaluate actions or states based on their values. *Exanna* incorporates values in state evaluation.

Organization

Section 5.2 discusses the related works. Section 5.3 details the *Exanna* framework. Section 5.4 describes a simulated pandemic scenario for evaluation and the results. Section 5.6 concludes our contributions and lists potential future works.

5.2 Related Work

Research on agents with values and rationales relates to our contributions.

5.2.1 Agents and Rationales

Hind et al. (2019) leverage existing supervised machine-learning techniques to generate rationales together with decisions without values involved and without exposing the inner details of the model. Whereas Hind et al. generate rationales based on the rationales in the training set, *Exanna* provides generalized and adaptive rationales based on context.

Georgara et al. (2022) propose an algorithm that wraps up any team formation algorithm to build justifications on why specific teams are formed, and some are not. Specifically, Georgara et al. build justifications based on contrastive explanations and by exploring what-if scenarios. A causal attribution explains why a behavior occur. In *Exanna*, we provide causal attribution of the selected action, precisely the premise, as rationales and wrap the rationales with values.

Wang et al. (2021) formulate rationales with the simplest subset of features with the proposed search algorithm. Specifically, the algorithm finds sufficient rationales by modifying the beam search algorithm and leveraging the tractability of expected predictions. The found set of features is sufficient as causal attribution for probabilistic solid guarantees on model behavior under observed data distribution. Contreras et al. (2022) propose a mirror model and assume a high understandability from performing similar to an observer's mental simulation. Contreras et al. apply deep Q-network and saliency maps in rationale generation, highlighting related input features as rationales. However, these works reveal each feature related to the model behavior.

Ajmeri et al. (2018) propose Poros, a framework that considers no values and shares full context as a rationale. Therefore, agents can make decisions from the perspective of others. In *Exanna*, agents *selectively* share information based on its and others' *values*.

5.2.2 Agents and Values

Ajmeri et al. (2020) present a framework that aggregates the value preferences of users to make ethically appropriate decisions. Lera-Leri et al. (2022) propose a method that considers a range of ethical principles from maximum utility to maximum fairness, for the aggregation of value systems instead of one single value. In *Exanna*, in addition to making decisions based on values, agents justify their behaviors and evaluate rationales based on its *values*.

Mosca and Such (2021) propose an agent that supports values in multiuser settings via generating optimal policy by considering the preferences and values of users. In addition, Mosca and Such justify solutions through contrastive explanations and positive answers. The causal attribution includes (1) a suggested action and the inputs of all users and (2) possible consequences from the user's preference. Whereas Mosca and Such generate explanations to present causal attribution with all the necessary information, *Exanna* further wraps rationales with values.

Agrawal et al. (2022) propose an agent that shares norms as causal attributions, and considers no values. Specifically, each agent evolves and learns rules of optimal behaviors with no values involved. Whereas norms define prevalent behavior in society, Agrawal et al. share learned rules as norms in homogeneous societies. However, *Exanna adaptively* shares learned rules that align with individual's *values* as rationales.

Montes and Sierra (2021) propose a methodology to synthesize parametric normative systems based on value promotion. Whereas Montes and Sierra focus on the design of moral norms synthesis and do not consider internal reasoning on norms, *Exanna* focuses on internal reasoning and justifying behavior based on values. Ogunniye and Kökciyan (2023) propose an ontology to represent the privacy domain which includes norms for social contexts, privacy preferences, and privacy policies. Ogunniye and Kökciyan introduce an argumentation-based dialogue to provide justifications during multi-party dialogues. In addition, the dialogue helps agents to reason about contextual norms and resolve privacy conflicts among agents. Di Scala and Yolum (2023) propose a new privacy agent for content concealment, equity of treatment, the collaboration of users, and the rationalization of actions. Specifically, Scala and Yolum provide textual output by considering outcomes and providing feedback to the user, such as a summary or detailed advice on possible actions to improve performance. Whereas these works claim that argumentation-based dialogue facilitates the exchange of rationales, *Exanna* provides a value-centered rationale for the made decision.

Table 5.1 summarizes the above comparisons emphasizing values and rationales.

5.3 Method

Exanna includes two components, rationale generation and evaluation, in addition to decision making. In this section, we first describe the schematics and payoff calculation with values. In the following subsections, we describe the decision making in *Exanna* along with the rationale components.

An *Exanna* agent acts and justifies its decisions based on values. An *Exanna* agent also observes other agents' actions, evaluates their rationales, updates its beliefs, and sanctions other agents based on the evaluation. With the rationale generation component, *Exanna* discloses private information based on values and causal attribution. Via evaluating with the rationale component, an *Exanna* agent mentally simulates other agents and responds with its values.

5.3.1 Schematics of a *Exanna* Agent

Algorithm 4 outlines an agent's decision-making loop, consisting of two phases through the agent's corresponding components: Rationale Generation and Evaluation. An *Exanna* agent has the following elements.

Decision rule is the mapping between an observation and a reasonable action, represented as if-then logic. A decision rule includes a premise and a consequence. The premise of a decision rule is a set of attribute-binding pairs. The consequent of a decision rule is an action to be taken when the premise holds. An example rule of Example 6 is

```
{InfectionRisk=No risk, InteractWith=Colleague} => Not Wear
```

Norm The expected behaviors or the behaviors of the majority in a group. When a majority applies the same decision rule, the rule becomes a norm. In *Exanna*, a norm uses the same if-then representation as a decision rule.

Sanction is the response to norm violation or satisfaction. A sanction can be a positive, negative, or neutral reaction from one agent to another.

Belief is the model or interpretation of the world, which is formed based on observations. We represent beliefs as b_t, indicating the belief from observation at time t. We store the beliefs as pairs of attributes and their bindings. Take Example 5 for instance. From Alice's perspective, the infection risk in the surroundings may be high because Bella has some symptoms. However, Bella believes the environment is safe based on her observation of Alice and her actual state—she has an allergy.

Goal defines the desired state that an agent wants to achieve. The outcome of a goal has a binary value, indicating whether the goal is achieved or not after performing the selected action

Action is the methodology to change the state and, therefore, approach the goals. We represent an action as a where $a \in A$ and A is the set of available actions.

Payoff refers to the outcome or result an agent receives in a given state after taking an action. The payoff is equal to reward r_t and feedback and sanction $saction_{partner}$ in our method

Context is the information that characterizes the situation of an entity. The subset of beliefs describes a context, which is represented as a set of attribute-binding pairs. An example of context is as below.

```
{InfectionRisk=No risk , Preference=Not Wear, InteractWith=Colleague
```

A context is given by observable attributes and nonobservable attributes in beliefs. Observable attributes are public and can be observed directly, e.g., the location where the agent is. However, nonobservable attributes are private to an agent and cannot be observed by other parties, e.g., another agent's preferences.

Values are context-specific as not all, but a subset of values is applicable within a context (Liscio et al. 2021). Values guide an individual's behaviors by changing the payoff function. v_i denotes the weight of one value in one value preference ($v_i \in V_{context}$) where $0 \leq v_i \leq 1$ and $\sum_{i=1}^{n} v_i = 1$.

Value preferences $V_{context}$ is a preference order over different values for one context. We store each value preference $V_{context}$ in a tuple where numbers add up to 1. We can treat each $\langle$V-in-context$\rangle$ as an attribute and store the corresponding preferences as its binding. For instance, an agent with value preferences $V = \{V_{pandemic} = \{v_{health} = 0.6, v_{privacy} = 0.4\}; V_{normal} = \{v_{health} = 0.4, v_{privacy} = 0.6\}\}$ indicates that the agent values health over privacy during a pandemic but the opposite in a normal context.

Preference refers to a subjective inclination for an option over other alternatives, which reflects the values of freedom in our settings. Whereas value preferences define the general motivational goals of an agent (Schwartz 2012), a preference is a specific inclination within a context. We represent a preference as p and $p \in O$ where O is the option space. In our simulation settings, $O = A$ where A is the action space.

Learning mechanism An *Exanna* agent learns from its experience with XCS, a rule-based learning algorithm that evolves rules from interactions. Specifically, the agent learns the best action based on the relationship between the observations and reward function. We will describe further details in the following subsections. Our reward function includes rewards r_t and sanctions $sanction_{partner}$ from others

We apply values in decision-making, rationale generation, and evaluating rationales. *Exanna* has two components to accommodate rationales from two perspectives: of the producer of a rationale for a decision and of the receiver of a rationale to reason about it. *Exanna* applies a rule-based learning algorithm to learn and evolve decision rules, which we describe further in the following subsection. Once an agent makes a decision, it generates a rationale based on learned decision rules for the chosen action with the rationale generation component. In addition, the agent wraps the rationale based on its and others' values. With the rationale evaluation component, an *Exanna* agent evaluates the rationale with an analogy upon receiving a rationale.

5.3.2 Payoff Calculation with Values

Utility theory studies how agents make decisions among multiple desirable alternatives (Edwards 1954). Specifically, each individual makes its decision to maximize expected utility based on their preferences. Whereas preferences define the tendency of an individual to make a subjective selection among alternatives, values define the important things to an individual. Although both values and preferences are context-specific, values may transcend contexts (Liscio et al. 2021).

Whereas agents get payoffs from their actions, we model them as $M_{individual}$ in Table 5.2. Each agent stores values in a tuple where each value maintains a corresponding $M_{individual}$. Since agents do not make decisions with single values but with tradeoffs among multiple related values, we aggregate value preferences when constructing payoff (Ajmeri et al. 2020). Below, f is the aggregated payoff with all corresponding values after selecting strategy Rx when the other player selects strategy Cy from $M_{individual}$.

$$f = \sum_{i}^{values} v_i \times r_{RxCy} \tag{5.1}$$

We model interactions as games with payoffs f from the aggregation of $M_{individual}$.

5.3.3 Decision Making

Figure 5.1 demonstrates the interaction between agents, specifically, the rationale giver and receiver. An actor agent selects an action based on its goal. Based on the chosen action, the actor agent can provide the rationale to the observer who witnesses its behavior. Upon receiving a rationale from the actor agent, the observer agent evaluates the rationale by making an analogous decision.

Algorithm 4 describes the pseudo-code of an agent's decision-making loop. An agent forms beliefs b_t about the world based on its observations in Line 4. Some beliefs may be observable, and some nonobservable by others. Agents may form different or even mutually inconsistent beliefs due to partial observability. Take Example 6 for instance, Bella chooses not to wear a mask since the infection risk is low in Bella's perspective. However, the infection risk may be high in Alice's perspective. The Q function in Line 6 and reward in Line 7 refers to the payoff calculation in Section 5.3.2. An agent's payoff is a weighted sum of payoffs corresponding to an agent's values. The Q function, in addition, includes feedback from others. In Line 6, the agent selects the action that gives the best payoff for b_t. If any to interact with, an agent builds the rationales based on b_t and the selected action at Line 9 with Algorithm 5. After performing the selected action, other agents that observe it and receive the rationales evaluate the rationales (Algorithm 6) with their context and give sanctions.

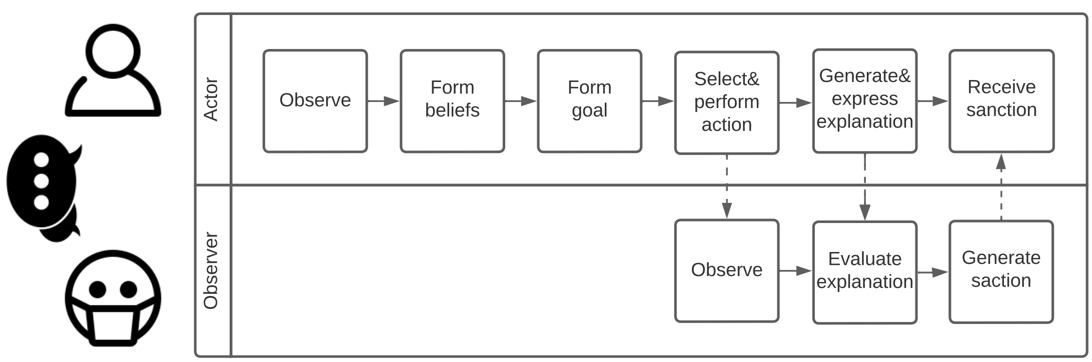


Figure 5.1: The interactions between *Exanna* agents. An agent forms its belief based on its observations. The beliefs are the agent's understanding of the world, which includes observable attributes in the context and attributes that are nonobservable by the other party. During the interaction, an agent makes its decision based on its goal and justifies the decision. Upon receiving the rationale, other agents evaluate and decide whether to accept it.

Algorithm 4: Decision-making for an *Exanna* agent.

```
Initialize agent (including value preferences V and other mental states);
Initialize rule-value function Q;
for t=1,T do
    Form beliefs b_t based on perceived state;
    Identify available actions A;
    With a probability ε select a random action a_actor ∈ A
      Otherwise select a_actor = argmax_a Q(b_t, a) ;
    Execute action a_actor and observe reward r_t;
    if Any observer agent pa then
        /* Generate rationales based on selected action and beliefs
           */
        Exp = GenRationale(b_t, a_actor);
        Send Exp to pa;
        Observe agent pa's action a_partner;
        if Receive rationales Exp_partner from agent pa then
            Update beliefs b_t based on Exp_partner;
        end
        /* Generate sanctions based on beliefs and given rationales
           */
        sanction_partner = EvalRationale(Exp_partner if any, a_partner, b_t);
    end
    /* Agents learn from reward and sanction                        */
    learn(b_t, a_actor, r_t + sanction_partner, b_{t+1});
end
```

5.3.4 Rationale Generation

Rule learning is the process of evolving rules from datasets or interactions. The basic form of rules is the if-then expression, e.g., if premise then consequent, where the consequent holds whenever the premise is true. A rule-based approach enables the flexibility and interpretability of norm implementation. We adapt XCS (Butz and Wilson 2000), a rule-based learning algorithm that utilizes a genetic algorithm and reinforcement learning, which evolves a set of rules or strategies based on payoffs or rewards produced by the proposed actions. An example rule of Example 6 is

```
{InfectionRisk=No risk, InteractWith=Colleague} => Not Wear.
```

The premise of a learned rule is a conjunction of attribute-binding pairs, e.g., {Infection-Risk=No risk, InteractWith=colleague}. The consequent of a learned rule is an action to be taken when the premise holds—in the above example, Not wear. With a weighted sum of payoffs, we incorporate values in decision-making where a substantial value casts a more significant effect on the final decision. XCS supports interpretability to humans with logical rules. Specifically, unlike other machine learning techniques, XCS generates a set of rules describing its decision in addition to the decision. A classifier rule sometimes called a rule or a classifier, is a pairing between a condition and an action. Each rule has associated (1) fitness indicating its suitability for continued use, (2) numerosity which indicates the number of instances of the rule in the rule set, (3) predicted reward indicating the expected reward if the rule applies, and (4) prediction error.

The overall process of XCS includes the following sub-processes.

- Matching: A process that matches the current context and all rules/classifiers to generate a match set. For instance, in Example 5, the match set for Bella may include (1) {InfectionRisk = Low} ⇒ Wear [fitness = 0.3], (2) {InfectionRisk = Low} ⇒ Not Wear [fitness = 0.7], (3) {OtherAgentType = Health} ⇒ Wear [fitness = 0.8], and (4){OtherAgentType = Health} ⇒ Not Wear [fitness = 0.2]. The fitness is based on the accuracy of each rule's reward prediction.

- Covering: A process that guarantees diversity via adding a random classifier whose conditions match the current context. For instance, adding {InfectionRisk = Low, Relationship = Friend} ⇒ Not Wear to the rule set.

- Action selection: XCS selects actions with pure exploration or pure exploitation with ε greedy. If not in exploration mode, this process returns the action with the highest fitness-weighted aggregation of reward.

$$fitness_a = \sum_{i}^{rule} fitness_i \times numerosity_i \times predicted\_reward_i \tag{5.2}$$

  where a $\in A$ and A is the action space. Rules represent all rules applied to the context and for action a. With the example and formula mentioned above, the agent would choose not to wear a mask due to $fitness_{notwear} > fitness_{wear}$.

- Formation of action set: The action set includes all classifiers that propose the chosen action based on the match set. For instance, {InfectionRisk = Low} ⇒ Not Wear, {OtherAgentType = Health} ⇒ Not Wear, and {InfectionRisk = Low, Relationship = Friend} ⇒ Not Wear.

- Updating classifier parameters (Urbanowicz and Browne 2017): An agent updates the rule parameters (e.g., accuracy and fitness) based on the received payoff. The following equation updates the predicted reward, where p is the predicted reward, β is the learning rate, and r is the received reward.

$$p \leftarrow p + \beta(r - p) \tag{5.3}$$

  The prediction error ε is updated with the following equation.

$$\varepsilon \leftarrow \varepsilon + \beta(|r - p| - \varepsilon) \tag{5.4}$$

  The fitness of a rule is based on its accuracy, which is inversely proportional to the prediction error. We update the accuracy *kappa* with the following formula.

$$\kappa = \begin{cases} 1 & if \;\; \varepsilon < \varepsilon_0 \\ \alpha(\frac{\varepsilon}{\varepsilon_0})^{-\nu} & otherwise \end{cases} \tag{5.5}$$

  where α is the scaling factor that raises a non-accurate rule to be close to an accurate rule. ε_0 is the threshold of prediction error below which the prediction error of a rule is assumed to be zero. ν defines how accuracy is related to prediction error and aims to help differentiate similar classifiers. For fitness calculation, we next calculate the relative accuracy κ' of each rule.

$$\kappa' = \frac{\kappa}{\sum_{cl \in [A]} \kappa_{cl}} \tag{5.6}$$

  where $[A]$ represents the corresponding action set. Finally, the fitness update of a rule is as follows.

$$F \leftarrow F + \beta(\kappa' - F) \tag{5.7}$$

  where F is the fitness of a rule.

- Subsumption: A process that replaces offspring rules with more general parent rules if it exists. Otherwise, save the offspring rules. Specifically, a more general rule yields a minor prediction error. For instance, if rule {InfectionRisk = Low} $\Rightarrow$ Not Wear has less prediction error than rule {InfectionRisk = Low, Raltionship = Friend} $\Rightarrow$ Not Wear, the former rule would replace the later rule and increases the numerosity.

- Deletion: Each action set has the same maximum number of rules. XCS removes the low-fitness rules.

Whereas context defines the information that can characterize the situation of an entity, not all context is necessary to share. For instance, in Example 6, sharing personal preference is unnecessary when both agents care about health. When generating a rationale, we first have XCS generate the base rationale from the learned rules and corresponding context. An example of a rationale for not wearing a mask is {InfectionRisk=No risk, Preference=Not Wear, InteractWith=colleague}. This rationale can be interpreted as no mask wearing when there is a no risk of infection, and the agent prefers not to wear a mask while interacting with a colleague in the office. Each agent keeps the rules it discovers and evolves in a rule set for decision-making. We further associate the factors in rules with values via wrapping the base rationale. Specifically, a learned rule may include observable and nonobservable information, leading to information overload and privacy leaks. Whereas values define what agents believe to be important, instead of sharing everything, we hide private information associated with what agents do not care about in the rationale. Following the example above, the agent who cares about health then adjusts its rationale to the colleague who cares about health to health-related causal attribution if it exists. For instance, no mask is required because of no risk of infection while interacting with a colleague in the office.

Algorithm 5 details the process of rationale generation. An agent first identifies rules associated with beliefs b_t at Line 2 and filters out rules not with the selected action at Line 3. The rationales are the aggregated rules (Line 4). To address our research questions, an agent only reveals hidden information associated with the values of agents involved in the interaction (Line 5 – 7). For instance, if an agent who cares about freedom interacts with one who cares about freedom, it will exclude the infection risk from the environment in its rationales. For each rationale, an agent evaluates the privacy loss as Line 8. The calculation of the privacy loss for each rationale is as below.

$$\text{Privacy Loss} = \frac{\text{number of otherwise nonobservable attributes shared in the rationale}}{\text{number of nonobservable attributes}} \tag{5.8}$$

In this work, we keep the privacy loss as a measure. However, privacy loss can be part of the rationale deliberation in future work.

5.3.5 Rationale Evaluation

Upon receiving a rationale from another agent, an agent first updates its beliefs based on the rationale. Specifically, the agent updates the beliefs of unobservable information from other party. With the previous example in rationale generation, the agent's colleague updates its beliefs

Algorithm 5: Rationale generation.

Input: beliefs b_t, Action a
Output: Rationale *Exp*

Function *GenRationale*:
  `/* Generate associated rules with beliefs` b_t `*/`
  Get match set *ms* with b_t;
  Generate action set from *ms* with *a*;
  Aggregate rules *Exp* associated with action set;
  if *values not involved* **then**
    remove factors in *Exp* that not related to presented values in b_t;
  end
  Privacy loss *pl* = number of nonobservable attributes shared in *Exp* / number of nonobservable attributes;
return

of the infection risk to no risk. In terms of XCS, the context update may change the associated rules and the corresponding match set. When evaluating a rationale, the agent's colleague makes an analogous decision based on the updated beliefs. If the selected action matches the observed action in the context, the agent receives no sanction. Otherwise, the agent violates its colleague's expectations and receives a negative sanction.

Algorithm 6 details the evaluation of given rationales. Upon receiving a rationale, an agent reasons over the rationale. Specifically, the agent first updates its beliefs b_t based on the rationale in Line 3, precisely the nonobservable context or beliefs of others. With the provided rationale, an agent checks if any applicable rules align with its rule sets in Line 4. The agent identifies associated rules from b_t and adds them to applicable rules in Line 5. In Line 8, the agent calculates the fitness for each available action for each applicable rule and keeps the best action for each rule. The agent accepts this rationale if any selected action matches the observed action.

5.4 Simulation

We evaluate *Exanna* via a simulated pandemic scenario based on Examples 5 and 6 where agents move to various places, interact with other agents, decide to wear or not wear a mask, and provide a justification for their actions. We implemented our pandemic environment using MASON (Luke et al. 2005), an agent-based simulation library in Java. Our focus is to investigate how different strategies of rationales change agents' behaviors. Agents in the simulation use XCS to evolve optimal strategies and maximally general classifiers, which justify the selection

Algorithm 6: Evaluating a rationale.

```
Input: Rationales Exp, Observed action a_partner, Beliefs b_t
Output: Decision d
Function EvalRationale:
    Initialize applicable rules ars;
    update b_t with hidden information in Exp;
    Add triggered match set from Exp to applicable rules ars;
    Add triggered match set from b_t to applicable rules ars;
    for rule in applicable rules ars do
        for act in possible actions do
            calculate fitness f_act;
        end
        Keep the act with best fitness;
    end
    if act contains a_partner then
        Decision d = accept;
    else
        Decision d = reject;
    end
return
```

of the strategies.

5.4.1 Scenario

The environment represents a society-based multiagent system with several locations and social circles. Our simulation involves a finite population of 200 agents with different social circles. The simulation has one park, one hospital, five homes, five offices, and five parties. Agents can move around and interact in five places (home, office, party park, and hospital). Each agent is native to one home, one office, and one party. Agents in the same home, office, or party share the same family, colleague, or friend relationship circle. In addition, each relationship circle involves 40 agents. In this simulated environment, time is represented in steps. Each agent moves to one location at each step and has a probability (50%) of interacting with one agent at the same location. Agents are more likely (75%) to move to places they are associated with when they move to home, office, and party, i.e., an agent is more likely to visit their own home than someone else's home.

Each agent forms its goal based on its value preferences. Specifically, each value in one context has a payoff matrix (Table 5.5 and 5.6), and the weighted sum of the payoff determines the goal (desired states). When an agent selects an action which does not align with its goal, it violates its goal. An agent optimizes its payoffs based on its value preferences. As Figure 5.1 illustrates, when an agent encounters another agent at the same location, it chooses an action based on its goal — whether to wear a mask. In addition, the agent justifies its behavior based on its beliefs in that context. For instance, the agent gives a rationale— {InfectionRisk=No risk, InteractWith=Colleague} —while not wearing a mask. The beliefs of an agent include observable attributes and attributes that are nonobservable by the other party. Due to the pandemic background, each agent receives a payoff according to the interaction location for action selection as in Table 5.3. Wearing a mask at a hospital during a pandemic is desirable. Location and value preferences determine the payoff an agent gives to itself. In addition, an agent also gives sanctions as feedback to others based on their actions. Upon observing the action and the corresponding agent's rationale in Figure 5.1, an agent evaluates the rationale with its beliefs and decides whether to accept it. If the rationale provides information not known to an agent earlier, the agent updates its beliefs based on the information. Take Example 5 for instance. Bella believes the infection risk in the context is low and chooses not to wear a mask. With the given rationale from Bella—{InfectionRisk=No risk, InteractWith=Colleague}—, Alice updates her beliefs on infection risk from high to low. With the updated beliefs, the agent makes its decision from the perspective of others. The agent accepts the rationale if the decision is the same as the observed behavior. Otherwise, the agent rejects the rationale. Furthermore, based on the decision on the rationale, the agent gives sanctions as feedback. The sanctions are based on the relationship circle. The closer the relationship is with someone, the more value is in the sanctions. Table 5.4 lists the sanctions associated with relationship circles.

We run each simulation 10 times, and each simulation lasts 30,000 steps. The values we consider include freedom and health. The value of freedom means agents would claim their free will and tend to follow their preferences.

5.4.2 Contextual Properties

Whereas agents have limited observations on the environment, the context includes the location (home, office, party, park, and hospital) where interactions occur, the relationship (family, friend, colleague, and stranger) with the observer, the subjective belief of infection risk of the environment, the personal preference on mask-wearing, and the types of observer agents. Due to the partial observation, agents would act based on their beliefs, whereas rationales enable belief updating. Table 5.7 lists the hidden contexts that are unobservable to others, which include infection risk, preference, and values.

5.4.3 Types of Societies

We define types of societies based on the rationale types. All societies include 50% of agents value health and 50% of agents value freedom. The value preferences of agents are as Table 5.8. All agents optimize their behaviors based on the weighted sum of payoffs from themselves and others.

Baseline 1: Share All Society In which agents share all information as rationales. Agents can make decision from the perspective of others.

Baseline 2: Share Decision Rules Society In which agents share their decision rules as rationales.

***Exanna*: Share Value-Aligned Rules Society** In which agents share their decision rules along with selective information that aligns with values as rationales.

5.4.4 Evaluation

To answer our research questions, we run simulations of share all, share decision rules, and *Exanna* societies. We propose the following hypotheses on social experience, goal adherence, privacy loss, and resolution.

$\mathbf{H}_{\textbf{Goal Adherence}}$. *Exanna* provides higher goal adherence than baseline societies

$\mathbf{H}_{\textbf{Conflict Resolution}}$. *Exanna* provides higher conflict resolution than baseline societies

$\mathbf{H}_{\textbf{Social Experience}}$. *Exanna* provides better social experience than baseline societies

$\mathbf{H}_{\textbf{Privacy Loss}}$. *Exanna* takes lower privacy loss compared to baseline societies

Below are the corresponding null hypotheses.

Exanna provides no difference in goal adherence compared to baseline societies

Exanna provides the same conflict resolution as baseline societies

Exanna provides no difference in social experience compared to baseline societies

Exanna takes the same privacy loss as baseline societies

To test the hypotheses, we measure the following metrics.

$\mathbf{M}_{\textbf{Goal Adherence}}$. The degree of adherence to each agent's goal. This metric ranges over [0, 1].

$\mathbf{M}_{\textbf{Conflict Resolution}}$. The percentage of conflict resolution in society. This metric ranges over [0, 100].

$\mathbf{M}_{\textbf{Social Experience}}$. The aggregation of payoff an agent receives for its behavior. This metric ranges over [-3, 3].

$\mathbf{M}_{\textbf{Privacy Loss}}$. The proportion of hidden information shared during an interaction. This metric ranges over [0, 1].

We compute the conflict resolution as follows.

$$\text{Conflict resolution} = \frac{\text{number of acceptance}}{\text{number of rationales}} \tag{5.9}$$

5.5 Results

We now discuss the results for our research questions. Table 5.9 summarizes the simulation results and the corresponding statistical analysis for our research questions. The metric row for each hypothesis in the table shows the numeric value of the metric during the simulation. All numeric values in Table 5.9 are rounded to 2 decimal places.

To evaluate these metrics, we conduct the independent t-test among different societies. We further measure effect size with Glass' Δ for different standard deviations among societies (Glass 1976; Grissom and Kim 2012). We adopt Cohen's (Cohen 1988) descriptors to interpret effect size where 0.2 indicate small, 0.5 indicate medium, and 0.8 indicate large effect. Specifically, an effect size less than 0.2 indicates negligible differences.

5.5.1 $\mathbf{H}_{\textbf{Goal Adherence}}$

$\mathrm{H}_{\text{Goal Adherence}}$ states that *Exanna* yield higher goal adherence across societies. The corresponding null hypothesis indicates no difference in goal adherence between societies. We measure the goal adherence in society to evaluate $\mathrm{H}_{\text{Goal Adherence}}$. Figure 5.2 compares the goal adherence for Share All, Share Decision Rules, and *Exanna* agent societies. Figure 5.3 details the payoff of different agent types in various agent societies. We find that *Exanna* has lower goal adherence (p-value < 0.01; $\Delta > 0.8$, indicating a large effect) than Share Decision Rules society. There is no significant difference (p-value > 0.05) between Share All society and *Exanna* society. However, the mean of goal adherence in the Share All society is higher than in the *Exanna* society.

Referring to Table 5.9, *Exanna* has lower goal adherence and yet, better social experience and resolution, which indicates that *Exanna* sacrifices some goals to resolve conflicts. The observed results follow our intuition that the more one can think from the logic of others, the more convincing the rationales are. The results of Share All and Share Decision Rules societies indicate that information overload is distracting when producing and reasoning with rationales. However, if an individual prefers to keep personal information and activity history private, the sacrifice of goal adherence is expected.

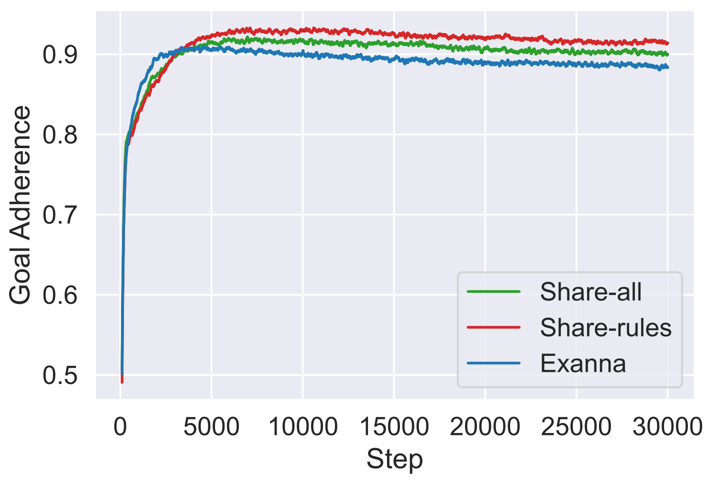
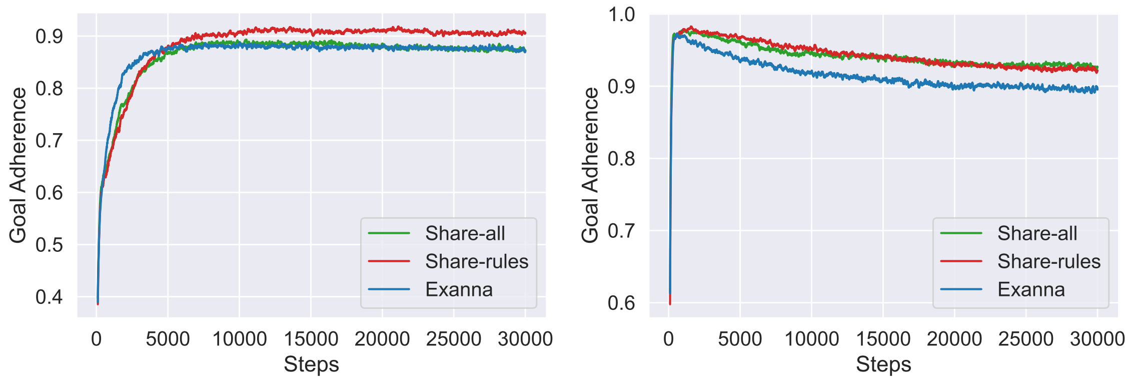


Figure 5.2: Comparing the goal adherence ($M_{\text{Goal Adherence}}$) in various agent societies. The *Exanna* agent society has lower total goal adherence (Glass' $\Delta > 0.8$; $p < 0.05$ for Share Decision Rules society but $p > 0.05$ for Share All society) than the baseline societies.

5.5.2 $H_{\text{Conflict Resolution}}$

$H_{\text{Conflict Resolution}}$ states that *Exanna* yields higher conflict resolution among societies. The corresponding null hypothesis indicates no difference in conflict resolution between societies. According to Table 5.9, *Exanna* has better conflict resolution (p-value < 0.001; $\Delta > 0.8$, indicating a large effect) than other societies. The analysis results reject the null hypothesis



(a) Goal adherence for health-freak agents. (b) Goal adherence for freedom-loving agents.

Figure 5.3: Comparing the goal adherence by agent types in various agent societies. The health-freak agents in Share Decision Rules society has higher goal adherence ($\bar{X} = 0.88$) than the Share All society ($\bar{X} = 0.862$ and *Exanna* society ($\bar{X} = 0.865$)). The freedom-loving agents in *Exanna* society has lower goal adherence ($\bar{X} = 0.91$) than the baseline societies ($\bar{X} = 0.94$).

that corresponds to $\text{H}_{\text{Conflict Resolution}}$. We find that agents in *Exanna* trade goal adherence with conflict resolution. Specifically, agents violate their original goals and gain higher acceptance in cases where providing rationales for following their heart does not convince others.

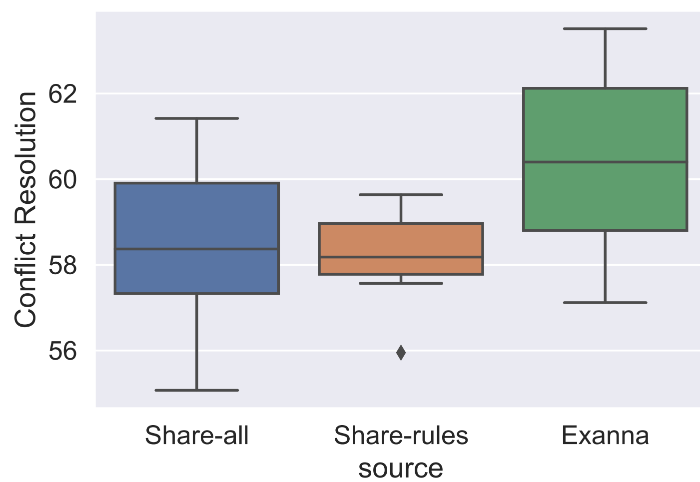


Figure 5.4: Comparing the resolution ($\text{M}_{\text{Conflict Resolution}}$) in various agent societies. The *Exanna* agent society has better resolution (Glass' $\Delta > 0.8$; $p < 0.001$) than the baseline societies.

5.5.3 $H_{\text{Social Experience}}$

$H_{\text{Social Experience}}$ states that *Exanna* provides better social experience across societies. The corresponding null hypothesis indicates no difference in social experience between societies. To evaluate $H_{\text{Social Experience}}$, we measure the overall payoffs of agents in society. An agent's payoff includes personal payoff from its action and the feedback from its interactor, if any. Figure 5.5 compares the social experience for Share All, Share Decision Rules, and *Exanna* agent societies. We find that *Exanna* yields better social experience (p-value < 0.001; $\Delta > 0.8$, indicating a large effect)) than other societies. Specifically, *Exanna* agents receive better feedback from other agents who receive the rationales. The analysis results reject the null hypothesis that corresponds to $H_{\text{Social Experience}}$.

After digging deeper into the results, we observe that *Exanna* agents receive more negative sanctions than other societies initially and soon learn to violate their goals to achieve better results. Figure 5.6 plots the payoff of the actor who selects an action and receives feedback from observers in various agent societies. Figure 5.7 shows the payoff from the observer who reacts to the actor's behavior in various agent societies. Whereas actors act and receive feedback from others, observers are agents who give feedback based on observed behaviors. The differences in actors' payoff among different agent societies are negligible. However, according to Figure 5.3 and Figure 5.7, agents in *Exanna* society have higher observer payoffs but deviate from the original goal more. That is, agents violate their goals and receive fewer negative sanctions. The goal violation brings better social experience and higher resolution than other baseline societies.

Intuitively, giving rationales alongside values would achieve a better social experience. In general, social experience includes privacy loss. Although we separate privacy loss from social experience for better measurement, the observed social experience plus privacy loss correspond to the expectation.

5.5.4 $H_{\text{Privacy Loss}}$

$H_{\text{Privacy Loss}}$ states that *Exanna* takes lower privacy loss among societies. The corresponding null hypothesis indicates no difference in privacy loss between societies. Figure 5.8 compares the privacy loss for Share All, Share Decision Rules, and *Exanna* agent societies. We find that the privacy loss yielded by *Exanna* to be better (p-value < 0.001; $\Delta > 0.8$, indicating a large effect) than other societies. The analysis results reject the null hypothesis that corresponds to $H_{\text{Privacy Loss}}$.

Although both Share Decision Rules and *Exanna* society share learned rules as rationales, *Exanna* further limits the shared private information to values that agents appraise. A rationale

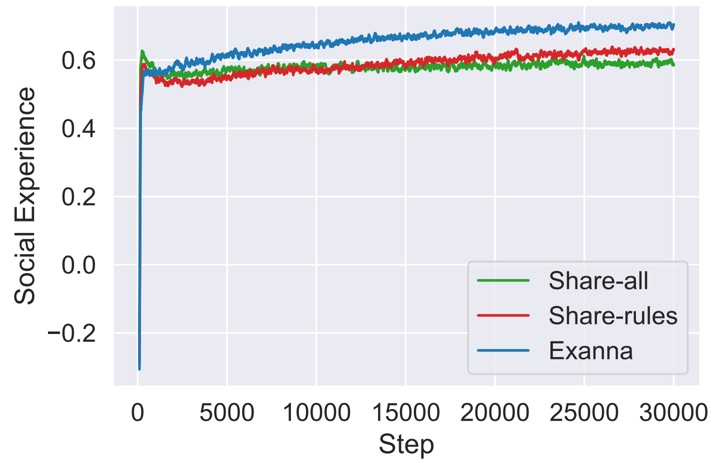
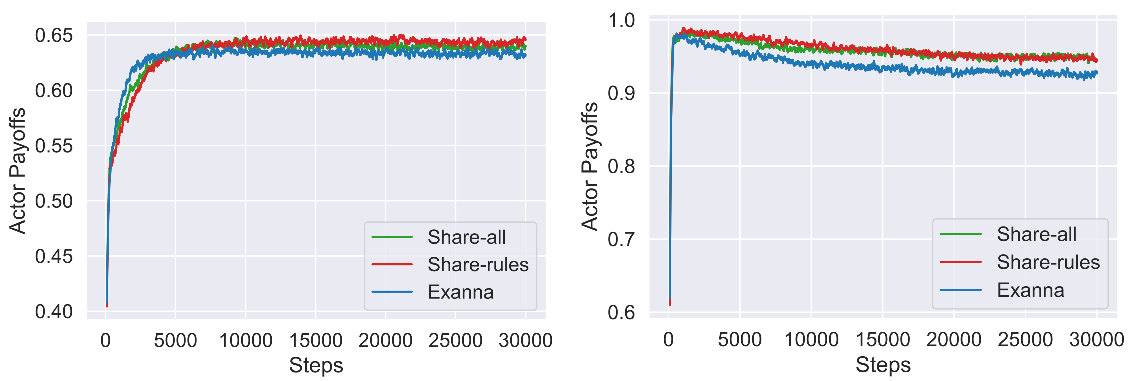


Figure 5.5: Comparing the social experience ($\mathrm{M}_{\text{Social Experience}}$) in various agent societies. The *Exanna* agent society has better social experience (Glass' $\Delta > 0.8$; $p < 0.001$) than the baseline societies.



(a) Actor payoff for health-freak agents.

(b) Actor payoff for freedom-loving agents.

Figure 5.6: Comparing the payoff of actors by agent types in various agent societies. Actors are agents who act and receive feedback from others. All societies have no significant difference in the actor payoff for health-freak agents. The freedom-loving agents in *Exanna* society has lower actor payoff ($\bar{X} = 0.94$) than the baseline societies ($\bar{X} = 0.96$).

stating causal attribution with the minimum private information and aligned with the values agents care about would be sufficient to explain behaviors. Intuitively, privacy loss could be part of the social experience. However, we separate them to observe the privacy loss when applying different rationalization strategies.

(a) Observer payoff for health-freak agents. (b) Observer payoff for freedom-loving agents.

Figure 5.7: Comparing the payoff from observers by agent types in various agent societies. Observers give feedback based on observed behaviors. The health-freak agents in *Exanna* society has better observer payoff ($\bar{X} = 0.02$) than Share All society ($\bar{X} = -0.05$) and Share Decision Rules society ($\bar{X} = -0.07$). The freedom-loving agents in *Exanna* society has better observer payoff ($\bar{X} = -0.28$) than Share All society ($\bar{X} = -0.38$) and Share Decision Rules society ($\bar{X} = -0.35$).

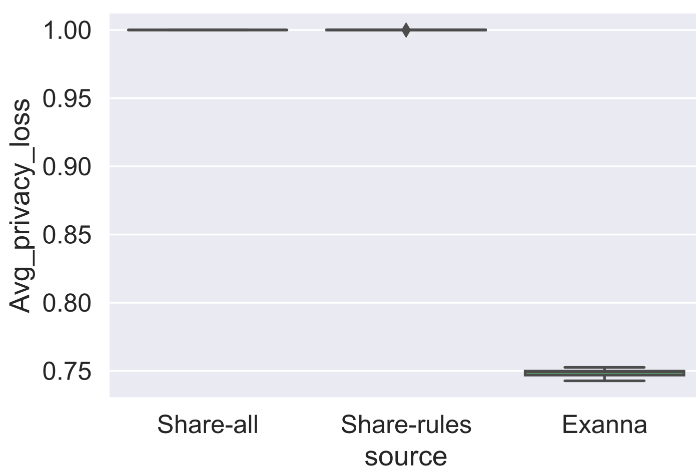


Figure 5.8: Comparing the privacy loss ($M_{\text{Privacy Loss}}$) in various agent societies. The *Exanna* agent society takes less privacy loss (Glass' $\Delta > 0.8$; $p < 0.001$) than the baseline societies.

5.5.5 Emerged Norm

A norm emerges when the proportion of agents adhering to a particular behavior surpasses a threshold. We consider 90% as the threshold (Delgado 2002). Table 5.10 lists the norms that emerge in the simulations.

5.6 Conclusions and Directions

Human-centered AI becomes a vital research domain when AI becomes part of our daily life. An agent must consider users and accommodate human factors in decision-making process. We present a framework for multiagent systems that reveals information in rationales based on agents' value.

Our results are consistent with the hypotheses except for goal adherence. First, value-aligned rationales trade some goal adherence for better social experience. Second, *Exanna* takes less privacy loss while maintaining higher resolution. Third, while value-aligned rationales wrap partial information, agents may learn to violate their goals while keeping their privacy. Specifically, agents who receive rejections from others learn to violate goals to some extent to maximize utility. Besides our simulation, *Exanna* can quickly adapt to other scenarios and values. With the updated fitness in XCS, *Exanna* is responsive to changes in multiagent systems.

5.6.1 Limitations and Threats to Validity

We made simplifying assumptions that, first, agents can identify other agents' types. Second, agent types indicate their values which guide their behaviors. These assumptions may not apply in all cases but are essential when interacting with other agents.

We simplified agents' value preferences to two values: health and freedom. This assumption may not apply to the real world, but it is crucial to demonstrate how value preferences shape behaviors. Specifically, a real-world scenario may include more intertwined values that are hard to quantify and tell which influences the decision more.

Whereas the required preliminary step for value-aligned AI is to identify values (Liscio et al. 2021, 2023) from stakeholders, we focus on demonstrating how incorporating values in decision-making and rationales shapes agent behaviors.

5.6.2 Future Directions

Whereas values diversify agents' decisions, this work suggests interesting directions. First, to include information cost in decision making. Different information may have different costs for agents, which may change agents' final decisions. For instance, sharing tax data and sharing interest pose significantly different costs. An agent may feel fine to share its interest but keep the tax data to itself. Second, to enable agents to decide what to share. In some cases, information suppression may be desirable. Third, to include rationales in decision-making instead of supplementary information. Having rationales as part of the decisions may increase the flexibility of an agent. Fourth, build an ontology to associate information with values, which we model as attributes. An ontology helps to model varied attributes or concepts and their intertwined relationships. Lastly, *Exanna* enables agents to act, justify their decisions, and reason about rationales based on an individual's decision rules. One future direction is to explore the emergence of shared norms and how agents pass and promote the norms.

Table 5.1: Summary of comparisons with related work. We compare works with respect to their application of values in decision-making and rationales (generation and evaluation). Rationale formation describes the representation of a rationale. The '–' notation indicates that no values are not applied in the case. Here, ✓ and ✗ signify that values are applied and not applied, respectively.

	Rationale	**Values applied in**		**Rationale formation**
		Decision	**Rationale**	
Wang et al. (2021)	✓	–	–	The prediction and a minimum subset of inputs but no information hiding
Contreras et al. (2022)	✓	–	–	Highlighted input features in deep Q-network but no information hiding
Ajmeri et al. (2020)	✗	✓	–	No rationales provided
Lera-Leri et al. (2022)	✗	✓	–	No rationales provided
Hind et al. (2019)	✓	✗	✗	Texts predicted via supervised learning, along with the predicted action
Agrawal et al. (2022)	✓	✗	✗	Norm as causal attribution but no information hiding
Ajmeri et al. (2018)	✓	✗	✗	Full context
Mosca and Such (2021)	✓	✓	✓	Suggested action based on the inputs from all users, along with the possible outcome of the user's preference as causal attribution but no information hiding
Ogunniye and Kökciyan (2023)	✓	✓	✓	A sequence of communications but no information hiding
Di Scala and Yolum (2023)	✓	✓	✓	Outcomes or advice based on complete information but no information hiding
Exanna	✓	✓	✓	Behavior rules (with information hiding) and alignment with values

Table 5.2: $M_{individual}$: Payoffs from agent interactions and from the environment.

Agent 1 \ Agent 2	C1	C2
R1	r_{R1C1}	r_{R1C2}
R2	r_{R2C1}	r_{R2C2}

Table 5.3: Actor's payoff associated with places. The numbers reflect general expectations of places.

Places	Wear	Not wear
Home	−0,25	0,25
Office	0,25	−0,25
Party	−0,25	0,25
Park	−0,5	0,5
Hospital	0,5	−0,5

Table 5.4: Feedback from an observer based on the relationship.

Social circle	Observer's response	
	Reject	Accept
Family	−1,00	1,00
Friend	−0,75	0,75
Coworker	−0,50	0,50
Stranger	−0,25	0,25

Table 5.5: Payoffs in terms of freedom depend on the agents' preferences.

Table 5.5(a) Payoffs corresponding to a preference for wearing a mask.

Agent 1 \ Agent 2	Wear	Not wear
Wear	1,0	1,0
Not wear	−1,0	−1,0

Table 5.5(b) Payoffs corresponding to a preference for not wearing a mask.

Agent 1 \ Agent 2	Wear	Not wear
Wear	−1,0	−1,0
Not wear	1,0	1,0

Table 5.6: Payoffs for the value of health. The numbers reflect how safe an agent feels.

Action \ Infection risk	No risk	High risk
Wear	0,0	1,0
Not wear	0,0	$-1{,}0$

Table 5.7: Nonobservable states that other agents cannot observe directly. The nonobservable states include others' mental states and information that is not revealed.

Type	Value
Beliefs on infection	No risk High risk
Preference	Wear Not wear
Values	Freedom Health

Table 5.8: Value preferences of agents.

Agents: Values	Freedom	Health
Freedom-loving	1,0	0,0
Health-freak	0,0	1,0

Table 5.9: Results: Comparing mean ($\bar{X}$) and standard deviation (σ) of social experience, resolution, goal adherence, and privacy loss in societies with Share All, Share Decision Rules, and *Exanna* agents.

		Share All	Share Decision Rules	*Exanna*
$M_{\text{Goal Adherence}}$	$\bar{X}$	0.9005	0.9138	0.8848
	σ	0.0099	0.0100	0.0253
	p-value	0.0831	< 0.01	–
	Δ	-1.5935	-2.8905	–
$M_{\text{Conflict resolution}}$	$\bar{X}$	0.5845	0.5819	0.6044
	σ	0.0198	0.0104	0.0211
	p-value	< 0.001	< 0.001	–
	Δ	1.8032	3.1065	–
$M_{\text{Social Experience}}$	$\bar{X}$	0.5910	0.6237	0.6994
	σ	0.0601	0.0244	0.0495
	p-value	< 0.001	< 0.001	–
	Δ	1.8032	3.1065	–
$M_{\text{Privacy Loss}}$	$\bar{X}$	1.0000	0,999	0.7486
	σ	0.0000	2.1130×10^{-5}	0.0029
	p-value	< 0.001	< 0.001	–
	Δ	∞	$-11\,896.5227$	–

Table 5.10: Emerged norms in agent societies. Common means the norms emerge in each agent society.

Society	Norm: Premise	Norm: Consequence
Common*	InfectionRisk=NO RISK; preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
	InfectionRisk=NO RISK; preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=HOSPITAL	WEAR
	InfectionRisk=RISK; preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
	InfectionRisk=RISK; preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=HOSPITAL	WEAR
Share All	InfectionRisk=NO RISK; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
Share Decision Rules	preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
Exanna	preference=NOT WEAR; InteractWith=COLLEAGUE; location=OFFICE	WEAR
	preference=NOT WEAR; InteractWith=COLLEAGUE; location=HOSPITAL	WEAR
	preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
	preference=NOT WEAR; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=HOSPITAL	WEAR
	OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
	OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=HOSPITAL	WEAR
	OberverAgentType=FREEDOM; InteractWith=COLLEAGUE; location=HOSPITAL	WEAR
	InfectionRisk=RISK; OberverAgentType=HEALTH; InteractWith=COLLEAGUE; location=OFFICE	WEAR
	InfectionRisk=NO RISK; OberverAgentType=HEALTH; InteractWith=COLLEAGUE;	WEAR

CHAPTER

6

CONCLUSION

This dissertation tackles the challenges of accommodating humans in the loop, i.e., emotion as sanctions, social signals as responses to norms, and human values as guidance of behaviors. We present a framework for a dynamic MAS that actively involves human participation. Our framework aims to accommodate humans in the loop and operates in dynamic environments. The human factors we target are expressed emotions, social signals, social value orientation, and values.

6.1 Answering the Research Questions

Noe, an agent framework that comprises emotional responses to the normative reasoning process. Emotion modeling in *Noe* enables the promotion of norm compliance and improvement of societal welfare.

Ness, an agent framework that models normative information from social signals to support norm emergence. Modeling soft signals such as hints or messages helps to avoid undesirable results and yields higher satisfaction than baseline agents despite requiring an equivalent amount of information.

Fleur, a framework that operationalizes the concept of Social Value Orientation (SVO). SVO provides agents with different preferences over resource allocations between themselves and others Aligning with SVO enables better social experience and robust norm emergence.

Exanna, a framework that incorporates values in decision-making, rationale generation, and reasoning over rationale. Constructing rationales based on agent values enhances social experience and conflict resolution at the cost of some goal adherence.

6.2 Future Directions

This work suggests numerous significant and captivating expansions. Our future directions includes delving further into comprehending the causal connections that exist between decisions and human factors.

For rationales of decisions, further research on how different costs of information influence decisions may enable more precise and reliable action suggestions and rationale construction. In addition, information suppression may be acceptable in some cases. Having agents decide what to share and when to share increases strategies' flexibility.

One interesting direction is investigating the relationship between social norms and different social signals. Whereas our work focuses on individual agent behaviors, expanding agent modeling from the micro to the macro level can help us understand and better define norms. Specifically, the micro level is the perspective of a single agent, and the macro level is of multiagent systems.

REFERENCES

Rishabh Agrawal, Nirav Ajmeri, and Munindar P. Singh. Socially intelligent genetic agents for the emergence of explicit norms. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 10–14, Vienna, July 2022. IJCAI. doi: 10.24963/ijcai.2022/2.

Stéphane Airiau, Sandip Sen, and Daniel Villatoro. Emergence of conventions through social learning. *Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 28(5):779–804, 2014. doi: 10.1007/s10458-013-9237-x.

Nirav Ajmeri, Pradeep K. Murukannaiah, Hui Guo, and Munindar P. Singh. Arnor: Modeling social intelligence via norms to engineer privacy-aware personal agents. In *Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 230–238, São Paulo, May 2017. IFAAMAS. doi: 10.5555/3091125.3091163.

Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Robust norm emergence by revealing and reasoning about context: Socially intelligent agents for enhancing privacy. In *Proceedings of the 27th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 28–34, Stockholm, July 2018. IJCAI. doi: 10.24963/ijcai.2018/4.

Nirav Ajmeri, Hui Guo, Pradeep K. Murukannaiah, and Munindar P. Singh. Elessar: Ethics in norm-aware agents. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS)*, pages 16–24, Auckland, May 2020. IFAAMAS. doi: 10.5555/3398761.3398769.

Bexy Alfonso Espinosa. *Agents with Affective Traits for Decision-Making in Complex Environments*. PhD thesis, Universitat Politècnica de València, 2017.

Giulia Andrighetto, Jordi Brandts, Rosaria Conte, Jordi Sabater-Mir, Hector Solaz, and Daniel Villatoro. Punish and voice: Punishment enhances cooperation when combined with norm-signalling. *PLOS ONE*, 8(6):1–8, 2013. doi: 10.1371/journal.pone.0064941.

Suwardi Annas, Muh Isbar Pratama, Muh Rifandi, Wahidah Sanusi, and Syafruddin Side. Stability analysis and numerical simulation of SEIR model for pandemic COVID-19 spread in Indonesia. *Chaos, Solitons & Fractals*, 139:110072, 2020. doi: 10.1016/j.chaos.2020.110072.

Estefania Argente, Elena Del Val, Daniel Perez-Garcia, and Vicente Botti. Normative emotional agents: A viewpoint paper. *IEEE Transactions on Affective Computing*, 13(3):1254–1273, 2022. doi: 10.1109/TAFFC.2020.3028512.

Gonul Ayci, Pınar Yolum, Arzucan Özgür, and Murat Şensoy. Explain to me: Towards understanding privacy decisions. In *Proceedings of the 22th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 2790—-2791, London, 2023. IFAAMAS. doi: 10.5555/3545946.3599079.

Lisa Feldman Barrett, Ralph Adolphs, Stacy Marsella, Aleix M. Martinez, and Seth D. Pollak. Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1):1–68, 2019.

Mathieu Bourgais, Patrick Taillandier, and Laurent Vercouter. Ben: An agent architecture for explainable and expressive behavior in social simulation. In *International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems*, pages 147–163, Cham, 2019. Springer. doi: 10.1007/978-3-030-30391-4_9.

Michael E. Bratman. *Intention, Plans, and Practical Reason*. Harvard University Press, Cambridge, Massachusetts, 1987.

Jan Broersen, Mehdi Dastani, Joris Hulstijn, Zisheng Huang, and Leendert van der Torre. The BOID architecture: Conflicts between beliefs, obligations, intentions and desires. In *Proceedings of the 5th International Conference on Autonomous Agents*, pages 9–16, New York, NY, USA, 2001. Association for Computing Machinery. doi: 10.1145/375735.375766.

Martin V. Butz and Stewart W. Wilson. An algorithmic description of XCS. In *Proceedings of the 3rd International Workshop on Learning Classifier Systems*, volume 1996 of *LNCS*, pages 253–272, Paris, France, 2000. Springer. doi: 10.1007/3-540-44640-0_15.

Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The quarterly journal of economics*, 117(3):817–869, 2002.

Amit K. Chopra and Munindar P. Singh. From social machines to social protocols: Software engineering foundations for sociotechnical systems. In *Proceedings of the 25th International World Wide Web Conference*, pages 903–914, Montréal, April 2016. ACM. doi: 10.1145/2872427.2883018.

Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, Hillsdale, New Jersey, 2nd edition, 1988. doi: 10.4324/9780203771587.

Victor Contreras, Michael Schumacher, and Davide Calvaresi. Integration of local and global features explanation with global rules extraction and generation tools. In *Explainable and Transparent AI and Multi-Agent Systems (EXTRAAMAS)*, volume 13283 of *LNCS*, pages 19–37, Virtual Conference, 2022. Springer. doi: 10.1007/978-3-031-15565-9_2.

Igor Conrado Alves de Lima, Luis Gustavo Nardin, and Jaime Simão Sichman. Gavel: A sanctioning enforcement framework. In *Proceedings of the 6th International Workshop on Engineering Multi-Agent Systems (EMAS)*, pages 225–241, Cham, 2019. Springer. doi: 10.1007/978-3-030-25693-7_12.

Jan de Mooij, Davide Dell'Anna, Parantapa Bhattacharya, Mehdi Dastani, Brian Logan, and Samarth Swarup. Quantifying the effects of norms on COVID-19 cases using an agent-based simulation. In *Proceedings of the The 22nd International Workshop on Multi-Agent-Based Simulation (MABS)*, pages 99–112, Cham, 2022. Springer International Publishing. doi: 10.1007/978-3-030-94548-0_8.

Carolyn H. Declerck and Sandy Bogaert. Social value orientation: Related to empathy and the ability to read the mind in the eyes. *The Journal of Social Psychology*, 148(6):711–726, 2008. doi: 10.3200/SOCP.148.6.711-726.

Jordi Delgado. Emergence of social conventions in complex networks. *Artificial Intelligence*, 141(1–2):171–185, 2002. doi: 10.1016/S0004-3702(02)00262-X.

Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. Runtime revision of norms and sanctions based on agent preferences. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1609–1617. IFAAMAS, 2019. doi: 10.5555/3306127.3331881.

Davide Dell'Anna, Mehdi Dastani, and Fabiano Dalpiaz. Runtime revision of sanctions in normative multi-agent systems. *JAAMAS*, 34(2):43:1–43:54, June 2020. doi: 10.1007/s10458-020-09465-8.

Daan Di Scala and Pınar Yolum. Paccart: Reinforcing trust in multiuser privacy agreement systems. In *Proceedings of the 22th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 2787–2789, London, 2023. IFAAMAS. doi: 10.5555/3545946.3599078.

Frank Dignum, Virginia Dignum, Paul Davidsson, Amineh Ghorbani, Mijke van der Hurk, Maarten Jensen, Christian Kammler, Fabian Lorig, Luis Gustavo Ludescher, Alexander Melchior, René Mellema, Cezara Pastrav, Loïs Vanhee, and Harko Verhagen. Analysing the combined health, social and economic impacts of the corovanvirus pandemic using agent-based social simulation. *Minds and Machines*, 30(2):177–194, 2020. doi: 10.1007/s11023-020-09527-6.

Ward Edwards. The theory of decision making. *Psychological Bulletin*, 51(4):380, 1954. doi: 10.1037/h0053870.

Christopher Frantz and Gabriella Pigozzi. Modeling norm dynamics in multiagent systems. *Journal of Applied Logics – IfCoLoG Journal of Logics and their Applications*, 5(2):491–564, 2018.

Athina Georgara, Juan A. Rodriguez Aguilar, and Carles Sierra. Building contrastive explanations for multi-agent team formation. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 516–524, Auckland, 2022. IFAAMAS. doi: 10.5555/3535850.3535909.

Adam Gerace. Internal and external attributions. In Virgil Zeigler-Hill and Todd K. Shackelford, editors, *Encyclopedia of Personality and Individual Differences*, pages 2328–2334. Springer, 2020. doi: 10.1007/978-3-319-24612-3_2301.

Gene V. Glass. Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5 (10):3–8, 1976. doi: 10.3102/0013189X005010003.

Donald W. Griesinger and James W. Livingston Jr. Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, 18(3):173–188, 1973. doi: 10.1002/bs.3830180305.

Robert J. Grissom and John J. Kim. *Effect Sizes for Research: Univariate and Multivariate Applications*. Routledge, Abingdon-on-Thames, 2012. doi: 10.4324/9780203803233.

Jayesh K. Gupta, Maxim Egorov, and Mykel Kochenderfer. Cooperative multi-agent control using deep reinforcement learning. In *Proceedings of the 9th Workshop on Adaptive Learning Agents workshop*, volume 10642 of *LNCS*, pages 66–83, São Paulo, 2017. Springer. doi: 10.1007/978-3-319-71682-4_5.

Jianye Hao, Jun Sun, Guangyong Chen, Zan Wang, Chao Yu, and Zhong Ming. Efficient and robust emergence of norms through heuristic collective learning. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 12(4):1–20, 2017. doi: 10.1145/3127498.

Michael Hind, Dennis Wei, Murray Campbell, Noel C.F. Codella, Amit Dhurandhar, Aleksandra Mojsilović, Karthikeyan Natesan Ramamurthy, and Kush R. Varshney. TED: Teaching AI to explain its decisions. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 123–129, Honolulu, 2019. ACM. doi: 10.1145/3306618.3314273.

Christopher D. Hollander and Annie S. Wu. The current state of normative agent-based systems. *Journal of Artificial Societies and Social Simulation*, 14(2):6, 2011. doi: 10.18564/jasss.1750.

Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. Revani: Revising and verifying normative specifications for privacy. *IEEE Intelligent Systems (IS)*, 31(5):8–15, September 2016. doi: 10.1109/MIS.2016.89.

Özgür Kafalı, Nirav Ajmeri, and Munindar P. Singh. DESEN: Specification of sociotechnical systems via patterns of regulation and control. *TOSEM*, 29(1):7:1–7:50, February 2020. doi: 10.1145/3365664.

Anup K. Kalia, Nirav Ajmeri, Kevin Chan, Jin-Hee Cho, Sibel Adalı, and Munindar P. Singh. The interplay of emotions and norms in multiagent systems. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 371–377, Macau, August 2019. IJCAI. doi: 10.24963/ijcai.2019/53.

Dacher Keltner and Jonathan Haidt. Social functions of emotions at four levels of analysis. *Cognition and Emotion*, 13(5):505–521, 1999. doi: 10.1080/026999399379168.

A. Kurtan and Pınar Yolum. Assisting humans in privacy management: An agent-based approach. *Autonomous Agents and Multi-Agent Systems*, 35(1):1–33, 2021. doi: 10.1007/s10458-020-09488-1.

Pat Langley. Explainable, normative, and justified agency. In *Proceedings of the 33nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 9775–9779, Honolulu, Hawaii, USA, 2019. AAAI Press. doi: 10.1609/aaai.v33i01.33019775.

Jeffrey W. Legro. Which norms matter? revisiting the "failure" of internationalism. *International Organization*, 51(1):31–63, 1997. doi: 10.1162/002081897550294.

Roger Lera-Leri, Filippo Bistaffa, Marc Serramia, Maite Lopez-Sanchez, and Juan Rodriguez-Aguilar. Towards pluralistic value alignment: Aggregating value systems through ℓp-regression. In *Proceedings of the 21st International Conference on Autonomous Agents* *and Multiagent Systems (AAMAS)*, pages 780–788, Auckland, 2022. IFAAMAS. doi: 10.5555/3535850.3535938.

Priel Levy and Nathan Griffiths. Convention emergence with congested resources. In *Proceedings of the 19th European Conference on Multi-Agent Systems*, pages 126–143, Cham, 2021. Springer. doi: 10.1007/978-3-030-82254-5_8.

Enrico Liscio, Michiel van der Meer, Luciano C. Siebert, Catholijn M. Jonker, Niek Mouter, and Pradeep K. Murukannaiah. Axies: Identifying and evaluating context-specific values. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 799–808, London, 2021. IFAAMAS. doi: 10.5555/3463952.3464048.

Enrico Liscio, Roger Lera-Leri, Filippo Bistaffa, Roel I.J. Dobbe, Catholijn M. Jonker, Maite Lopez-Sanchez, Juan A. Rodriguez-Aguilar, and Pradeep K. Murukannaiah. Value inference in sociotechnical systems. In *Proceedings of the 22th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1774–1780, London, 2023. IFAAMAS. doi: 10.5555/3545946.3598838.

Jamie Lopez Bernal, Nick Andrews, Charlotte Gower, Eileen Gallagher, Ruth Simmons, Simon Thelwall, Julia Stowe, Elise Tessier, Natalie Groves, Gavin Dabrera, Richard Myers, Colin N.J. Campbell, Gayatri Amirthalingam, Matt Edmunds, Maria Zambon, Kevin E. Brown, Susan Hopkins, Meera Chand, and Mary Ramsay. Effectiveness of COVID-19 vaccines against the B. 1.617. 2 (Delta) variant. *New England Journal of Medicine*, 385(7):585–594, 2021. doi: 10.1056/NEJMc2113090.

Maite Lopez-Sanchez, Marc Serramia, Juan A Rodriguez-Aguilar, Javier Morales, and Michael Wooldridge. Automating decision making to help establish norm-based regulations. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1613–1615, São Paulo, May 2017. IFAAMAS. doi: 10.5555/3091125.3091380.

Sean Luke, Claudio Cioffi-Revilla, Liviu Panait, Keith Sullivan, and Gabriel Balan. MASON: A multiagent simulation environment. *Simulation: Transactions of the Society for Modeling and Simulation International*, 81(7):517–527, July 2005.

Javier Marín-Morales, Juan Luis Higuera-Trujillo, Alberto Greco, Jaime Guixeres, Carmen Llinares, Enzo Pasquale Scilingo, Mariano Alcañiz, and Gaetano Valenza. Affective computing in virtual reality: Emotion recognition from brain and heartbeat dynamics using wearable sensors. *Scientific Reports*, 8(1):1–15, 2018. doi: 10.1038/s41598-018-32063-4.

Ofir Marom and Benjamin Rosman. Belief reward shaping in reinforcement learning. In *Proceedings of the 32nd Conference on Artificial Intelligence (AAAI)*, pages 3762–3769, New Orleans, 2018. AAAI Press. doi: 10.5555/3504035.3504496.

Stacy Marsella, Jonathan Gratch, and Paolo Petta. Computational models of emotion. In Klaus R. Scherer, Tanja Banziger, and Etienne Roesch, editors, *Blueprint for Affective Computing*, chapter 1.2, pages 21–41. Oxford University Press, 2010.

Stacy C. Marsella and Jonathan Gratch. EMA: A process model of appraisal dynamics. *Cognitive Systems Research*, 10(1):70–90, 2009. doi: 10.1016/j.cogsys.2008.03.005.

David Masad and Jacqueline Kazil. MESA: An agent-based modeling framework. In *Proceedings of the 14th PYTHON in Science Conference*, pages 53–60, Austin, 2015. Springer.

Mehdi Mashayekhi, Nirav Ajmeri, George F. List, and Munindar P. Singh. Prosocial norm emergence in multiagent systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 17(1–2):3:1–3:24, June 2022. doi: 10.1145/3540202.

Kevin R. McKee, Ian M. Gemp, Brian McWilliams, Edgar A. Duéñez-Guzmán, Edward Hughes, and Joel Z. Leibo. Social diversity and social preferences in mixed-motive reinforcement learning. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 869–877, Auckland, 2020. IFAAMAS. doi: 10.5555/3398761.3398863.

Stanley Milgram, Hilary J. Liberty, Raymond Toledo, and Joyce Wackenhut. Response to intrusion into waiting lines. *Journal of Personality and Social Psychology*, 51(4):683, 1986. doi: 10.1037/0022-3514.51.4.683.

Tim Miller. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, 267:1–38, 2019. doi: 10.1016/j.artint.2018.07.007.

Thomas M. Moerland, Joost Broekens, and Catholijn M. Jonker. Emotion in reinforcement learning agents and robots: A survey. *Machine Learning*, 107(2):443–480, 2018. doi: 10.1007/s10994-017-5666-0.

Nieves Montes and Carles Sierra. Value-guided synthesis of parametric normative systems. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 907–915, London, 2021. IFAAMAS. doi: 10.5555/3463952.3464060.

Javier Morales, Michael J. Wooldridge, Juan A. Rodríguez-Aguilar, and Maite López-Sánchez. Off-line synthesis of evolutionarily stable normative systems. *Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 32(5):635–671, 2018. doi: 10.1007/s10458-018-9390-3.

Andreasa Morris-Martin, Marina De Vos, and Julian Padget. Norm emergence in multiagent systems: A viewpoint paper. *Autonomous Agents and Multi-Agent Systems (JAAMAS)*, 33(6): 706–749, 2019. doi: 10.1007/s10458-019-09422-0.

Andreasa Morris-Martin, Marina De Vos, Julian Padget, and Oliver Ray. Agent-directed runtime norm synthesis. In *Proceedings of the 22th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 2271–2279, London, 2023. IFAAMAS. doi: 10.5555/3545946.3598905.

Francesca Mosca and Jose M. Such. ELVIRA: An explainable agent for value and utility-driven multiuser privacy. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 916–924, London, May 2021. IFAAMAS. doi: 10.5555/3463952.3464061.

Partha Mukherjee, Sandip Sen, and Stéphane Airiau. Norm emergence under constrained interactions in diverse societies. In *Proceedings of the 7th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 779–786, Estoril, 2008. IFAAMAS. doi: 10.5555/1402298.1402332.

Ryan O. Murphy and Kurt A. Ackermann. Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1): 13–41, 2014. doi: 10.1177/1088868313501745.

Pradeep K. Murukannaiah, Nirav Ajmeri, Catholijn M. Jonker, and Munindar P. Singh. New foundations of ethical multiagent systems. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1706–1710, Auckland, May 2020. IFAAMAS. doi: 10.5555/3398761.3398958. Blue Sky Ideas Track.

Luis G. Nardin, Tina Balke-Visser, Nirav Ajmeri, Anup K. Kalia, Jaime S. Sichman, and Munindar P. Singh. Classifying sanctions and designing a conceptual sanctioning process model for socio-technical systems. *The Knowledge Engineering Review (KER)*, 31(2):142–166, March 2016. doi: 10.1017/S0269888916000023.

Pablo Noriega, Harko Verhagen, Julian Padget, and Mark d'Inverno. Ethical online AI systems through conscientious design. *IEEE Internet Computing*, 25(6):58–64, 2021. doi: 10.1109/MIC.2021.3098324.

Gideon Ogunniye and Nadin Kökciyan. Contextual integrity for argumentation-based privacy reasoning. In *Proceedings of the 22th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 2253–2261, London, 2023. IFAAMAS. doi: 10.5555/3545946.3598903.

Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, New York, 1988. doi: 10.1017/CBO9780511571299.

Michael Pernpeintner. Toward a self-learning governance loop for competitive multi-attribute MAS. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems AAMAS*, pages 1619–1621, London, 2021. IFAAMAS. doi: 10.5555/3463952.3464179.

Piero Poletti, Marcello Tirani, Danilo Cereda, Filippo Trentini, Giorgio Guzzetta, Giuliana Sabatino, Valentina Marziano, Ambra Castrofino, Francesca Grosso, Gabriele Del Castillo, Raffaella Piccarreta, ATS Lombardy COVID-19 Task Force, Aida Andreassi, Alessia Melegaro, Maria Gramegna, Marco Ajelli, and Stefano Merler. Probability of symptoms and critical disease after SARS-CoV-2 infection. *arXiv*, 2006.08471:1–10, 2020. URL `https://arxiv.org/abs/2006.08471`.

Anand S. Rao and Michael P. Georgeff. Modeling rational agents within a BDI-architecture. In *Proceedings of the 2nd International Conference on Principles of Knowledge Representation and Reasoning*, pages 473–484, San Francisco, 1991. Morgan Kaufmann. doi: 10.5555/3087158.3087205.

John Realpe-Gómez, Giulia Andrighetto, Luis Gustavo Nardin, and Javier Antonio Montoya. Balancing selfishness and norm conformity can explain human behavior in large-scale prisoner's dilemma games and can poise human groups near criticality. *Physical Review E.*, 97(4):042321, 2018. doi: 10.1103/PhysRevE.97.042321.

Rudolph J. Rummel. Understanding conflict and war: Vol. 1: The dynamic psychological field. *Sage Publications*, 1975.

Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, New Jersey, 3 edition, 2010.

Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7(1):21–54, 2011. doi: 10.3233/MGS-2011-0167.

Bastin Tony Roy Savarimuthu, Stephen Cranefield, Maryam A. Purvis, and Martin K. Purvis. Obligation norm identification in agent societies. *Journal of Artificial Societies and Social Simulation*, 13(4):3, 2010. doi: 10.18564/jasss.1659.

Shalom H. Schwartz. An overview of the schwartz theory of basic values. *Online readings in Psychology and Culture*, 2(1):2307–0919, 2012. doi: 10.9707/2307-0919.1116.

Norbert Schwarz. Emotion, cognition, and decision making. *Cognition and Emotion*, 14(4): 433–440, 2000. doi: 10.1080/026999300402745.

Marc Serramia, Maite López-Sánchez, Juan A. Rodríguez-Aguilar, Manel Rodríguez, Michael Wooldridge, Javier Morales, and Carlos Ansótegui. Moral values in norm decision making. In *Proceedings of the 17th Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*, pages 1294–1302, Stockholm, July 2018. IFAAMAS. doi: 10.5555/3237383.3237891.

Herbert A. Simon. *The New Science of Management Decision*. Harper & Brothers, 1960. doi: 10.1037/13978-000.

Herbert A. Simon. Motivational and emotional controls of cognition. *Psychological Review*, 74 (1):29–39, 1967. doi: 10.1037/h0024127.

Amika M. Singh and Munindar P. Singh. Norm deviation in multiagent systems: A foundation for responsible autonomy. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1–9, Macau, August 2023. IJCAI.

Munindar P. Singh. *Multiagent Systems: A Theoretical Framework for Intentions, Know-How, and Communications*. Number 799 in LNCS. Springer, Heidelberg, 1994. doi: 10.1007/BFb0030531. URL `http://www.csc.ncsu.edu/faculty/mpsingh/books/MAS/`.

Munindar P. Singh. Norms as a basis for governing sociotechnical systems. *TIST*, 5(1): 21:1–21:23, December 2013. doi: 10.1145/2542182.2542203.

Paul Slovic. The construction of preference. *American Psychologist*, 50(5):364, 1995. doi: 10.1037/0003-066X.50.5.364.

Aron Szekely, Francesca Lipari, Alberto Antonioni, Mario Paolucci, Angel Sánchez, Luca Tummolini, and Giulia Andrighetto. Evidence from a long-term experiment that collective risks change social norms and promote cooperation. *Nature Communications*, 12:5452:1–5452:7, September 2021. doi: 10.1038/s41467-021-25734-w.

Mark G. Thompson, Jefferey L. Burgess, Allison L. Naleway, Harmony L. Tyner, Sarang K. Yoon, Jennifer Meece, Lauren E.W. Olsho, Alberto J. Caban-Martinez, Ashley Fowlkes, Karen Lutrick, Jennifer L. Kuntz, Kayan Dunnigan, Marilyn J. Odean, Kurt T. Hegmann, Elisha Stefanski, Laura J. Edwards, Natasha Schaefer-Solle, Lauren Grant, Katherine Ellingson, . . . , and Manjusha Gaglani. Interim estimates of vaccine effectiveness of BNT162b2 and mRNA-1273 COVID-19 vaccines in preventing SARS-CoV-2 infection among health care personnel, first responders, and other essential and frontline workers—eight US locations, December 2020–March 2021. *Morbidity and Mortality Weekly Report*, 70(13):495, 2021. doi: 10.1001/jamanetworkopen.2020.16382.

Myrthe L. Tielman, Catholijn M. Jonker, and M. Birna Van Riemsdijk. Deriving norms from actions, values and context. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 2223–2225, 2019. doi: 10.5555/3306127.3332065.

Stavroula Tsirogianni, Gordon Sammut, and Eri Park. *Social Values and Good Living*. Springer Netherlands, 2014. doi: 10.1007/978-94-007-0753-5_3666.

Sz-Ting Tzeng, Nirav Ajmeri, and Munindar P. Singh. Noe: Norms emergence and robustness based on emotions in multiagent systems. In *Proceedings of the International Workshop on Coordination, Organizations, Institutions, Norms and Ethics for Governance of Multi-Agent Systems (COINE)*, volume 13239 of *LNCS*, pages 62–77, London, May 2021. Springer. doi: 10.1007/978-3-031-16617-4_5.

Ryan J. Urbanowicz and Will N. Browne. *Introduction to Learning Classifier Systems*. Springer Briefs in Intelligent Systems. Springer, New York, 2017. doi: 10.1007/978-3-662-55007-6.

Paul A. M. Van Lange. The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology*, 77(2): 337–349, August 1999. doi: 10.1037/0022-3514.77.2.337.

Henricus J.E. Verhagen. *Norm autonomous agents*. PhD thesis, Stockholm Universitet, 2000.

Christian von Scheve, Daniel Moldt, Julia Fix, and Rolf von Luede. My agents love to conform: Norms and emotion in the micro-macro link. *Computational and Mathematical Organization Theory*, 12(2–3):81–100, 2006. doi: 10.1007/s10588-006-9538-6.

Eric Wang, Pasha Khosravi, and Guy Van den Broeck. Probabilistic sufficient explanations. In *Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 3082–3088, Montreal, August 2021. IJCAI. doi: 10.24963/ijcai.2021/424.

Christopher J.C.H. Watkins and Peter Dayan. Q-learning. *Machine Learning*, 8(3–4):279–292, 1992. doi: 10.1007/BF00992698.

Michael Winikoff, Galina Sidorenko, Virginia Dignum, and Frank Dignum. Why bad coffee? explaining BDI agent behaviour with valuings. *Artificial Intelligence*, 300:103554, 2021. doi: 10.1016/j.artint.2021.103554.

Jessica Woodgate and Nirav Ajmeri. Macro ethics for governing equitable sociotechnical systems. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 1824–1828, Online, May 2022. IFAAMAS. doi: 10.5555/3535850.3536118. Blue Sky Ideas Track.

Yang Wu and Laura E Schulz. Understanding social display rules: Using one person's emotional expressions to infer the desires of another. *Child Development*, 91(5):1786–1799, 2020. doi: 10.1111/cdev.13346.

Yang Wu, Chris L. Baker, Joshua B. Tenenbaum, and Laura E Schulz. Rational inference of beliefs and desires from emotional expressions. *Cognitive Science*, 42(3):850–884, 2018. doi: 10.1111/cogs.12548.

Chayu Yang and Jin Wang. A mathematical model for the novel coronavirus epidemic in Wuhan, China. *Mathematical Biosciences and Engineering MBE*, 17(3):2708, 2020. doi: 10.3934/mbe.2020148.

APPENDICES